Hydrophobic Gating and Functional Annotation of Ion Channels

# Hydrophobic Gating and Functional Annotation of Ion Channels




**Gianni Klesse**

Department of Physics

University of Oxford

A thesis submitted for the degree of Doctor of Philosophy.

St Cross College — Trinity 2019

Copyright © 2019 Gianni Klesse

All rights reserved.

The *viva voce* examination was held on 21.11.2019 by Prof Philip Biggin (University of Oxford) and Dr Marco Cecchini (University of Strasbourg). A copy of this thesis has been deposited in the Bodleian Library, Oxford.

*Thousands have lived without love, not one without water.*

– W. H. Auden

# Abstract

Ion channels are integral membrane proteins that facilitate the permeation of ions through lipid bilayers. By means of a molecular mechanism known as hydrophobic gating, these proteins can close their conduction pathway to ion flow without the need for steric occlusion. Hydrophobic gates pose a challenge in the analysis of novel ion channel structures, as they cannot be detected by measuring the physical dimensions of a channel pore. This thesis therefore presents the Channel Annotation Package (CHAP), an open-source software tool for the functional annotation of ion channel structures based on molecular dynamics (MD) simulation and capable of reliably indentifying hydrophobic gates. The cation-selective 5-$HT_3$ receptor and the anion-selective glycine receptor served as test cases to demonstrate the utility of CHAP. The functional status of several recent structures of these two channels was determined through MD simulation and hydrophobic gates were identified in both receptors. Furthermore, a high-throughput survey of ion channel structures enabled by CHAP revealed that hydrophobic gates occur in $\sim$20 % of the ion channel structural proteome. Next-generation polarisable force fields were used to explore the importance of electronic polarisation in MD simulations of ion channels. Water molecules passing a hydrophobic gate experienced a significant reduction of their dipole moment, mirroring the behaviour of water in a liquid-vapour phase transition. Polarisable ions interacted more strongly with uncharged protein residues than simulations employing traditional additive force fields suggest. Incorporating polarisation effects will thus be an important aspect of future computational studies of ion channel structure and function. The impact of a transmembrane potential on the hydrophobic gate in the 5-$HT_3$ receptor was studied through atomistic MD simulations. Electrowetting occured only at supraphysiological voltages, while at lower field strengths pore de-wetting effectively prevented ion permeation. The hydration response of the channel could be quantitatively described by a thermodynamic model that includes the influence of the charged amino acid residues of the channel protein. Electric field effects are unlikely to play an important role for hydrophobic gating in biological ion channels, but may potentially be exploited in the design of gated artificial nanopores.

# Acknowledgements

First and foremost, I would like to thank my supervisors, Prof Stephen Tucker and Prof Mark Sansom. In many ways the past four years felt like a great voyage of exploration. For much of this journey, my supervisors gave me a great amount of leeway to decide where to take my research. They let me be my own captain and provisioned me with all the resources I needed. At the same time, Stephen and Mark served as my map and compass and always pointed me in the right direction. By giving me the freedom to shape my own project, they gave me the opportunity to grow into an independent researcher. But without their guidance I would surely have suffered shipwreck. For all this, I am truly and deeply grateful.

Out of the many members of the SBCB, who helped me over the course of my DPhil studies, more than a few deserve explicit mention. Countless rewarding discussions with Dr Charlotte Lynch were a great source of insight and encouragement for me. Her astute comments sharpened my reasoning and her meticulous proofreading was a blessing for this thesis. I also owe a debt of gratitude to Shanlin Rao, who always patiently shared her knowledge of Biochemistry with me. Her efforts in testing the first prototype of CHAP were instrumental in creating a robust and user-friendly programme. It was a privilege to work with them both!

My research also benefitted greatly from the knowledge of Dr Irfan Alibay, whose expertise in high performance computing was invaluable to the success of this project. With Sarah-beth Amos I shared common interests in Bayesian statistics, in wine, and in dogs. Our discussions were always helpful and many times served as an antidote to the torments of thesis writing. In addition, I am grateful to Dr Prafulla Aryal and Dr Viwan Jarerattanachat for teaching me the ropes of molecular dynamics simulations, to Prof Phillip Stansfeld, for his assistance with biomembrane simulations, to Dr Keith Cassidy, for his `Tcl`-wizardry, to Dr Tom Newport, for thoughts on software, to Dr Jan Domanski, for thoughts on science, and to Dr Jemma Trick, for laying the ground upon which this project was built. Appreciation is due to all postdocs, students, and academic visitors who made the SBCB a fantastic place to work.

Valuable guidance also came from outside the Tucker and Sansom labs. My college advisor, Dr Jürgen Brem, is to be thanked for his insights and his uplifting comments. I am also thankful to Dr Richard Bradshaw, for helpful discussions that cleared the ground for my work with the AMOEBA force field, to Prof Chris-Kriton Skylaris, Dr Jacek Dziedzic, and Dr James Womack, for introducing me to linearly scaling density functional theory, and to Simone Rubinacci, for his advice on software development. To Prof Ramin Golestanian and Dr Mira Zorkot I am grateful for their supervision during my DTC rotation project. Gratitude is also owed to the Biochemistry IT team, Richard Bryan and John Elder, for their adapt guardianship over my research data.

Fortunately, while I was working on water, I did not have to live without love. I am deeply indebted to my parents, Brigitte and Michael Klesse, who have been an endless source of support and encouragement. My sister Jana also deserves my appreciation for always lending an ear to my woes. To my partner Tiny I am tremendously grateful for constantly being supportive and caring — no matter the distance.

Many more people made my time in Oxford memorable and I would like to thank them all for their friendship and support. I am especially grateful to Bartosz, whose perseverance continues to be an inspiration to me, and to Lennard, for his advice and his courage. Thanks is also due to Christine and David, who made St Cross a home away from home, and to Roberta, who did the same at Venneit Close. João and Robert are owed thanks for being excellent travel companions and wonderful hosts. Anton, Sandra, Michael, Sherin, Robert, Katja, and Simon accompanied me through countless great adventures, and so too did Nina, Andreas, and Florian — I am grateful to all of them. For always brimming with infectious mirth, Alicia and Augustin have my lasting admiration. Finally, thanks is due to my friends in the junge Deutsche Physikalische Gesellschaft, who have inspired and guided me.

Financial support from the Engineering and Physical Sciences Research Council and the St Cross College Travel and Research Fund is gratefully acknowledged.

Gianni Klesse
Oxford, October 2019

# Contents

**List of Figures** xiii

**List of Tables** xv

**List of Publications** xvii

**1 Introduction** 1

1.1 Ion Channels . . . 1

1.1.1 Biophysical Properties and Physiological Roles . . . 3

1.1.2 Relating Structure and Function . . . 5

1.2 Hydrophobic Gating . . . 7

1.2.1 The Hydrophobic Gating Mechanism . . . 7

1.2.2 Hydrophobic Gates in Biological Ion Channels . . . 10

1.2.3 Experimental Support for Hydrophobic Gating . . . 12

1.2.4 Thermodynamic Theory of Hydrophobic Gating . . . 14

1.3 Molecular Dynamics for Functional Annotation . . . 17

1.4 Thesis Outline . . . 19

**2 Theory and Methods** 23

2.1 Classical Molecular Dynamics . . . 23

2.1.1 Pairwise Additive Force Fields . . . 25

2.1.2 Water Models . . . 30

2.1.3 Molecular Dynamics Algorithms . . . 33

2.2 Polarisation Effects in Classical Molecular Dynamics . . . 42

2.2.1 The Classical Drude Oscillator . . . 42

2.2.2 Polarisable Atomic Multipoles . . . 49

2.3 Umbrella Sampling . . . 57

2.4 Continuum Electrostatics . . . 59

**3 The Channel Annotation Package** 63

3.1 Workflow for Functional Annotation . . . 65

3.2 Implementation of CHAP . . . 68
3.2.1 Pore Geometry and Centre Line . . . 71
3.2.2 Water Density and Free Energy . . . 75
3.2.3 Hydrophobicity Profiles . . . 80
3.3 Further Implementation and Simulation Details . . . 85
3.3.1 Molecular Dynamics Simulation Protocol . . . 85
3.3.2 B-Spline Curves . . . 86
3.4 Summary and Conclusions . . . 89

**4 Hydrophobic Gates in Ion Channels 91**
4.1 Methods . . . 93
4.1.1 Simulations of the 5-$HT_3$ Receptor . . . 93
4.1.2 Simulations of the Glycine Receptor . . . 93
4.1.3 High Throughput Analysis of Ion Channel Structures . . . 94
4.2 Results . . . 95
4.2.1 Hydrophobic Gating in the 5-$HT_3$ Receptor . . . 95
4.2.2 Hydrophobic Gating in the Glycine Receptor . . . 110
4.2.3 Prevalence of Hydrophobic Gates in Biological Ion Channels . . . 119
4.3 Discussion . . . 121
4.4 Summary and Conclusions . . . 124

**5 The Influence of Induced Polarisation 125**
5.1 Methods . . . 129
5.1.1 Equilibrium Molecular Dynamics Simulation . . . 129
5.1.2 Umbrella Sampling . . . 134
5.1.3 Analysis . . . 136
5.2 Results . . . 139
5.2.1 Validation of Polarisable Simulation Protocol . . . 140
5.2.2 Influence of Polarisation on Water Behaviour . . . 144
5.2.3 Influence of Polarisation on Ion Conduction . . . 161
5.3 Discussion . . . 176
5.4 Summary and Conclusions . . . 179

**6 Electrowetting of Hydrophobic Gates 183**
6.1 Methods . . . 186
6.1.1 Molecular Dynamics Simulation . . . 186
6.1.2 Poisson-Boltzman Electrostatics . . . 189

6.1.3 Analysis . . . . . 190
6.2 Results . . . . . 195
6.2.1 Liquid-Vapour Equilibrium Under an Electric Field . . . . . 195
6.2.2 Intrinsic Electric Field of the 5-$HT_3$ Receptor . . . . . 203
6.2.3 Ionic Currents . . . . . 208
6.2.4 Kinetics of Liquid-Vapour Transitions . . . . . 212
6.3 Discussion . . . . . 218
6.4 Summary and Conclusions . . . . . 221

**7 Summary and Conclusions** **223**
7.1 The Channel Annotation Package . . . . . 223
7.2 Hydrophobic Gates in Biological Ion Channels . . . . . 225
7.3 The Influence of Induced Polarisation . . . . . 226
7.4 Electrowetting of Hydrophobic Gates . . . . . 227
7.5 Conclusion . . . . . 228

**A Polarisable Force Field Benchmark** **229**

**Bibliography** **231**

# List of Figures


1.1 Molecular Transport Through Biological Membranes . . . . . 2
1.2 Ion Channels Classified by Gating Type . . . . . 4
1.3 Ion Channel Structures in the Protein Data Bank . . . . . 6
1.4 Liquid-Vapour Oscillations and Hydrophobic Gating . . . . . 8
1.5 Hydrophobic Gating in the TWIK-1 Potassium Channel . . . . . 13
1.6 Kelvin Equation for Cylindrical Capillaries . . . . . 15
1.7 Definition of Contact Angle . . . . . 16

2.1 Interaction Types of Molecular Dynamics Force Fields . . . . . 26
2.2 Functional Form of Common Force Field Terms . . . . . 28
2.3 Geometry of Common Additive Water Models . . . . . 31
2.4 Molecular Dynamics Simulation Workflow . . . . . 34
2.5 Geometry of the CHARMM Drude Water Models . . . . . 44
2.6 Geometry of the AMOEBA Water Models . . . . . 54

3.1 Physical Dimensions of the 5-$HT_3$ Receptor Pore . . . . . 64
3.2 Functional Annotation Workflow . . . . . 66
3.3 Identification of Hydrophobic Gates Through Equilibrium MD . . . . . 68
3.4 Workflow Diagram of Data Processing in CHAP . . . . . 70
3.5 Pore Geometry Calculation . . . . . 72
3.6 Convergence of Nelder-Mead Optimisation . . . . . 73
3.7 Comparison of Narrowest Constriction Radii . . . . . 74
3.8 Determination of Water Density . . . . . 77
3.9 Influence of Bandwidth Parameter . . . . . 79
3.10 Hydrophobicity Profile of Pore-Lining Residues . . . . . 82
3.11 Hydrophobicity Scales Available in CHAP . . . . . 84

4.1 Structural Architecture of the 5-$HT_3$ Receptor . . . . . 98
4.2 5-$HT_3$ Receptor Pore . . . . . 99
4.3 5-$HT_3$ Receptor Radius and Hydrophobicity Profiles . . . . . 100
4.4 De-Wetting of 5-$HT_3$ Receptor Pore . . . . . 103

4.5 Influence of 5-$HT_3$ Receptor Structural Model . . . 105
4.6 Influence of Positional Restraints . . . 108
4.7 Structural Architecture of the Glycine Receptor . . . 111
4.8 Glycine Receptor Pore . . . 113
4.9 De-Wetting of Glycine Receptor Pore . . . 115
4.10 Influence of Glycine Receptor Structural Model . . . 118
4.11 Pore Dewetting Across the Ion Channel Structural Proteome . . . 120

5.1 Macroscopic Properties of Common Water Models . . . 126
5.2 Equilibration of the Polarisable Multipole Simulation System . . . 141
5.3 Equilibration of Drude Oscillator Simulation System . . . 143
5.4 Protein Position in Polarisable Bilayer . . . 144
5.5 Comparison of Pore Hydration Using Polarisable Water Models . . . 146
5.6 Comparison of Pore Hydration Using Additive Water Models . . . 148
5.7 Hydration Equilibrium in the 5-$HT_3$ Receptor Pore . . . 150
5.8 Water Dipole Moment Profile in 5-$HT_3$ Receptor . . . 154
5.9 Components of Water Dipole Moment . . . 157
5.10 Sensitivity of Dipole Moment Profile to Structural Model . . . 158
5.11 Schematic Summary of Water Dipole Moment Behaviour . . . 160
5.12 Single Ion Potential of Mean Force Profile . . . 162
5.13 Convergence of Umbrella Sampling . . . 165
5.14 Ionic Hydrations Shells . . . 169
5.15 Hydration Number Profile for $Na^+$ Ions . . . 171
5.16 Hydration Number Profile for $Cl^-$ Ions . . . 174

6.1 Electrowetting of Hydrophobic Gates . . . 184
6.2 Electrowetting of the 5-$HT_3$ Receptor . . . 197
6.3 Beckstein Openness in Dependence of External Electric Field . . . 198
6.4 Posterior Probability of Bayesian Model Parameters . . . 201
6.5 Electrostatic Potential of the 5-$HT_3$ Receptor . . . 205
6.6 Parameter Dependence of Electrostatic Potential Profile . . . 206
6.7 Pore-Facing Side Chain Orientation . . . 208
6.8 Cumulative Charge Flow Through the 5-$HT_3$ Receptor . . . 209
6.9 Ionic Current Through the 5-$HT_3$ Receptor . . . 211
6.10 Kinetics of Liquid-Vapour Transitions . . . 214
6.11 Ion Dwell Time . . . 217

# List of Tables

2.1 Parameters of Common Additive Water Models . . . . . 32
2.2 Parameters of the CHARMM Drude Water Models . . . . . 45
2.3 Parameters of the AMOEBA Water Models . . . . . 55

4.1 5-$HT_3$ Receptor Structures . . . . . 92
4.2 Glycine Receptor Structures . . . . . 94

5.1 Bulk Hydration Properties of $Na^+$ Ions . . . . . 168
5.2 Bulk Hydration Properties of $Cl^-$ Ions . . . . . 172

6.1 Bayesian Model Parameters . . . . . 202

A.1 Polarisable Force Field Performance Benchmark . . . . . 229

# List of Publications

The work presented in this thesis has contributed to the following publications:

- **G. Klesse**, S. Rao, M. S. P. Sansom, and S. J. Tucker. "CHAP: A Versatile Tool for the Structural and Functional Annotation of Ion Channel Pores". In: *Journal of Molecular Biology* 431.17 (2019), pp. 3353-3365.
- S. Rao, **G. Klesse**, P. J. Stansfeld, S. J. Tucker, and M. S. P. Sansom. "A Heuristic Derived From Analysis of the Ion Channel Structural Proteome Permits the Rapid Identification of Hydrophobic Gates". In: *Proceedings of the National Academy of Sciences* 116.28 (2019), pp. 13989-13995.
- S. Rao, C. I. Lynch, **G. Klesse**, G. E. Oakley, P. J. Stansfeld, S. J. Tucker, and M. S. P. Sansom. "Water and Hydrophobic Gates in Ion Channels and Nanopores". In: *Faraday Discussions* 209 (2018), pp. 231-247.
- P. S. Miller, S. Masiulis, T. Malinauskas, A. Kotecha, S. Rao, S. Chavali, L. De Colibus, E. Pardon, S. Hannan, S. Scott, Z. Sun, G. Frenz, **G. Klesse**, S. Li, J. M. Diprose, C. A. Siebert, R. M. Esnouf, F. DiMaio, S. J. Tucker, T. G. Smart, J. Steyaert, M. M. Babu, M. S. P. Sansom, J. T. Huiskonen, and A. R. Aricescu. "Heteromeric GABAA Receptor Structures in Positively-Modulated Active States". In: *bioRxiv* (2018), p. 338343.
- S. Rao, **G. Klesse**, P. J. Stansfeld, S. J. Tucker, and M. S. P. Sansom. "A BEST Example of Channel Structure Annotation by Molecular Simulation". In: *Channels* 11.4 (2017), pp. 347-353.
- J. L. Trick, S. Chelvaniththilan, **G. Klesse**, P. Aryal, E. J. Wallace, S. J. Tucker, and M. S. P. Sansom. "Functional Annotation of Ion Channel Structures by Molecular Simulation". In: *Structure* 24.12 (2016), pp. 2207-2216.

This is my thesis. There are
many like it, but this one is mine.

*(After Jerry Gustav Hasford)*

# 1

# Introduction

## 1.1 Ion Channels

Biological organisms are typically segregated into distinct aqueous compartments in order to create favourable conditions for the chemical reactions that underlie their metabolism and facilitate their reproduction.[1] In prokaryotes, the cytosolic region inside the cell is separated from the outside environment by a plasma membrane made up of lipids and a cell wall consisting of proteins and carbohydrates. Gram negative bacteria have an additional outer membrane separating the cell wall from the extracellular space. By contrast, eukaryotic cells typically have a single plasma membrane, but are internally compartmentalised into various organelles, such as mitochondria or the nucleus, which are separated from the cytosol through their own membranes.

Biological membranes are typically composed of amphipathic lipid molecules that self-assemble into bilayer structures. The central region of such lipid bilayers is highly hydrophobic and thus impermeable to polar molecules and ions (Figure 1.1). This allows biological membranes, which are typically only a few nanometres thin, to act as efficient capacitors. Cells exploit this for the purpose of energy storage by creating ionic concentration gradients between the extracellular and cytosolic domains, which gives rise to a net electric field across the membrane.

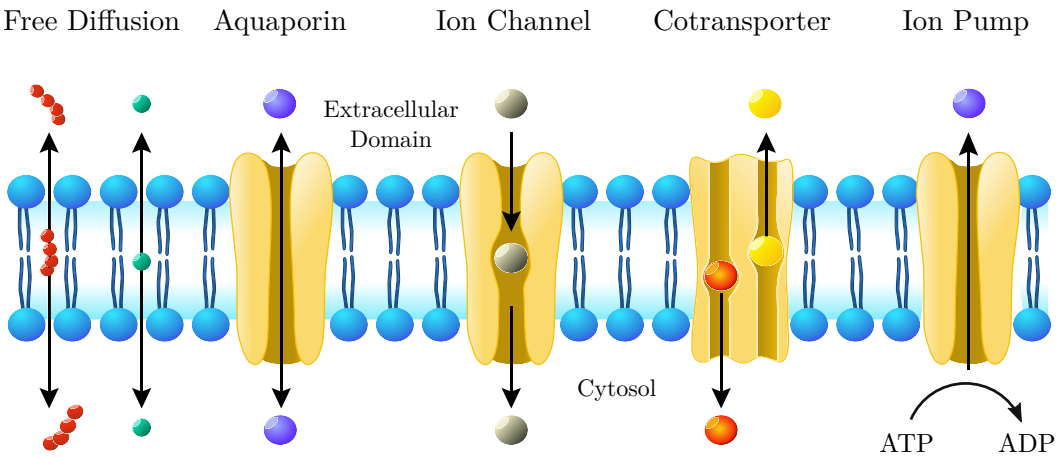



**Figure 1.1:** Molecular transport through biological membranes. While hydrophobic molecules like oxygen or lipid-derived hormones can freely diffuse through a lipid bilayer, the transport of polar and charged species is facilitated by specialised integral membrane proteins. Aquaporins enable the fast transport of water, while ion channels control the diffusion of selected ionic species along electro-chemical gradients. Ion pumps actively transport ions by expending energy derived from ATP hydrolysis. Active ion transport against an electro-chemical gradient can also be accomplished by cotransporters, which either transport ions of equal charge in opposite directions (antiporters, shown) or ions with opposite charges in the same direction (symporters, not shown). Copyright 2016 iStock.com/ttsz.

These vital electro-chemical gradients are controlled by two classes of integral membrane proteins: (active) ion transporters and (passive) ion channels. Transporters expend energy to facilitate the motion of ions against their respective concentration gradient. In a process termed primary active transport, the energy required to accomplish this is derived from the hydrolysis of adenosine triphosphate (ATP). Proteins employing this mechanism are referred to as ion pumps.

Alternatively, in a process called secondary active transport, so-called cotransporters exploit the energy stored in an existing transmembrane potential. Cotransporters are proteins that transport two different ions simultaneously, one against and the other along its respective concentration gradient. If the two ions carry opposite charges and move in the same direction, the protein is referred to as symporter; if they carry the same charge and move in opposite directions, the protein is called an antiporter.

By contrast, ion channels facilitate passive transport along an existing electro-chemical

gradient, which does not directly require an external source of energy. While transporters can carry ions at a rate of up to $\sim 1 \times 10^3\,\mathrm{s}^{-1}$, ion channels are characterised by a much higher conductivity and can achieve transport rates from $\sim 1 \times 10^6\,\mathrm{s}^{-1}$ up to $\sim 1 \times 10^9\,\mathrm{s}^{-1}$, close to the rate of ionic diffusion in bulk electrolyte.[2] This type of fast ion transport is exploited by physiological processes that rely on rapid electrical signalling, such as the initiation and propagation of action potentials in the nervous system or the contraction of muscle fibres. Ion channels achieve these high transport rates by providing an aqueous pathway through the hydrophobic core of the lipid bilayer, through which ions can freely diffuse in response to the electro-chemical gradient.

### 1.1.1 Biophysical Properties and Physiological Roles

However, ion channels are not merely static pores in the membrane, but instead exhibit a rich dynamical behaviour. In response to external stimuli, channel proteins can undergo allosteric transitions between conformations that either permit or prohibit the diffusive flow of ions. This is known as ion channel gating and the conductive and non-conductive conformations are referred to as the open and closed states of the channel.

What kind of stimulus will induce a gating transition depends on the physiological role of the ion channel (Figure 1.2). Voltage-gated gated channels react to changes in the transmembrane potential and facilitate the transport of ions that underlies the electrical excitability of nerve and muscle cells.[3] Ligand-gated channels sense the extracellular binding of small molecules, typically neurotransmitters, and are responsible for the transduction of chemical into electrical signals.[4] Mechano-sensitive channels gate in response to variations in pressure or stress and are involved in the sensation of touch,[5] the maintenance of osmotic pressure balance in prokaryotes,[6] and mechanoperception in plants.[7]

While most channels only have a single open state with a uniquely defined single channel conductance, multiple non-conductive states are common.[8] For example, voltage-

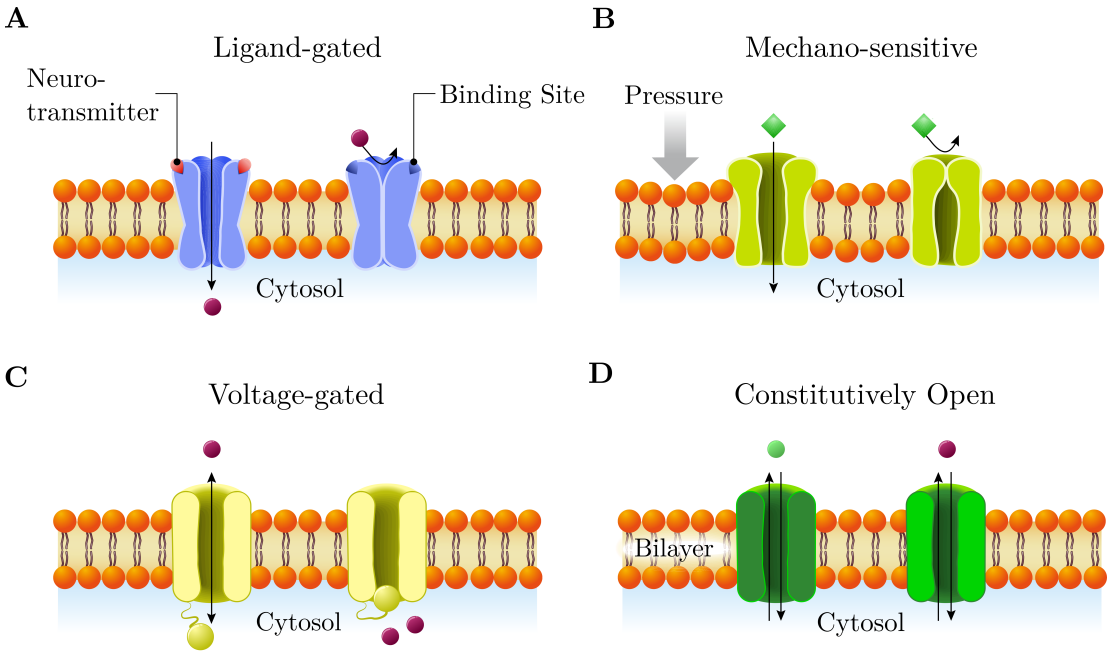



**Figure 1.2:** Ion channels gate in response to a wide variety of stimuli depending on their physiological role. (A) Ligand-gated channels are activated by the binding of small molecules and are responsible for neurotransmitter sensing. (B) Mechanosensitive channels respond to changes in pressure or stress. (C) Voltage-gated channels enable fast electrical signalling by propagating action potentials. (D) Some ion channels are not gated and give rise to small leakage or background currents. Copyright 2016 iStock.com/ttsz.

gated $Na^+$ channels typically become inactivated a short time after opening in order to permit a repolarisation of the membrane and to prevent the backpropagation of action potentials. Similarly, ligand-gated channels may assume a desensitised conformation and become impermeable to ions in the continued presence of the relevant agonist.

In addition to their ability to gate, ion channels are also often characterised by a high degree of selectivity, i.e. they can be permeable only to certain ionic species, but not to others.[9] Most channels discriminate at least between cations and anions, but many have also evolved the remarkable capability to selectively conduct one specific ion species. For example, potassium channels conduct $K^+$ ions at a rate ~100 times faster than $Na^+$ ions, even though $Na^+$ has a significantly smaller ionic radius.[8]

Due to their involvement in a wide range of physiological processes, defects in ion chan-

nels are the cause of a large number of diseases.[8] These so-called channelopathies have been related to mutations in over 60 ion channel encoding genes in *homo sapiens*.[10] They include cardiac arrhythmias,[11,12] neurological disorders like ataxia[13] and epilepsy,[14] as well as neonatal diabetes[15] and cystic fibrosis.[16] While only ∼1 % of all human genes encode ion channels,[17] approximately 13.4 % of the currently available drugs are targeted at ion channel proteins, which thus constitute the second largest family of drug targets.[18,19] In addition, ion channels are a main target of general anaesthetic agents.[20]

### 1.1.2 Relating Structure and Function

Their physiological relevance and their importance in drug development have placed ion channels in the focus of structural biologists. Structural biology aims to elucidate the molecular mechanisms underlying ion channel function by identifying the protein domains responsible for ligand-binding, voltage sensing, and stress activation, as well as the structural elements of the channel pore that act as selectivity filter and as channel gate.

The first high resolution structures of ion channels were predominantly determined by X-ray crystallography and, since the publication of the structure of the KcsA potassium channel in a seminal study in 1998,[21,22] the number of channel structures deposited in the Protein Data Bank (PDB) has grown steadily (Figure 1.3). Insights gleaned from these early structures were instrumental in identifying the basic patterns of ion channel architecture and the development of mechanistic models for their function.[23] However, due to their primarily hydrophobic surface and the presence of detergent used to solubilise the protein, membrane proteins are notoriously challenging to crystallise,[24,25] which has long been an obstacle in X-ray crystallography studies of these proteins.

Over the past decade, improvements in detector technologies and image processing algorithms have allowed cryo-electron microscopy (cryo-EM) experiments to achieve atomic resolution.[27] Since cryo-EM does not require the crystallisation of proteins, this

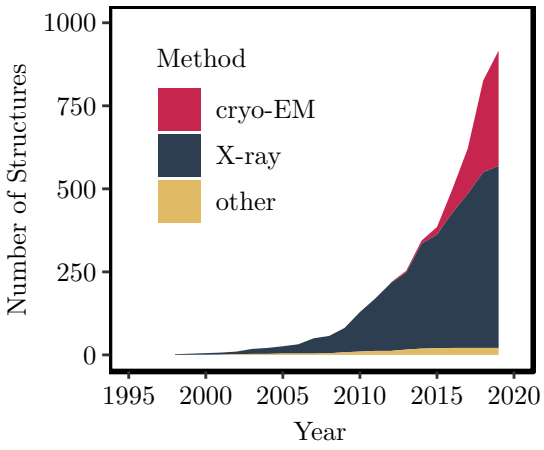



**Figure 1.3:** The number of ion channel structures deposited in the Protein Data Bank is growing exponentially and may well exceed 1000 by the end of the decade. An increasing proportion of ion channel structures originates from cryo-EM experiments. Figure adapted from Klesse et al. [26] under a creative commons (CC BY 4.0) licence with additional data retrieved from the PDB on 1 August 2019.

has led to a revolution in the structural biology of membrane proteins[28] and has yielded a wealth of new structures.[29] Unsurprisingly, these advancements in cryo-EM are also reflected in a burst of novel ion channel structures published since the middle of this decade (Figure 1.3) and have enabled the discovery of channel structures of unprecedented size and complexity, such as those of the ryanodine receptor[30] or the mechanosensitive Piezo1 channel.[31]

In tandem with these structural investigations, computer simulations have been used to study the dynamical behaviour of ion channels and to derive the biophysical principles that connect the structure of an ion channel to its function.[32–35] A central objective of these studies is to determine how ion channels gate, i.e. to provide a physical explanation for the molecular mechanism that prevents the permeation of ions in the closed state of a channel and to identify which conformational rearrangements lead to the opening of a gate.

In the most simple cases, gate closing is based on the physical obstruction of the permeation pathway and the blockage of ion flow through steric clashes with pore-lining side chains. However, in many channels the gating mechanism turns out to be more sophisticated. In particular, it has been shown that ion channels can exploit the hydrophobic effect to create an energetic barrier to ion permeation even in pores that

are sufficiently large to accommodate a hydrated ion. This effect is now known as hydrophobic gating and has been demonstrated to be an important process in several ion channels and non-biological pores.[36–38]

The following section will examine the hydrophobic gating mechanism in detail. It comprises a survey of biological ion channels, in which hydrophobic gates have been observed, and a review of the experimental evidence in support of hydrophobic gating. Following this, the thermodynamic theory that forms the basis of a quantitative understanding of hydrophobic gating is discussed.

## 1.2 Hydrophobic Gating

The concept of hydrophobic gating originates from simulation studies of model nanopores designed to mimic ion channels, which showed that water confined in the interior of such pores exhibits a complex phase behaviour depending on the physico-chemical character of the pore lining.[39–41] In nanopores with a hydrophobic inner surface, water may undergo alternating evaporation and condensation events and the pore thus stochastically transitions between a "dry" and a "wet" state (Figure 1.4A). By contrast, water confined by polar surfaces exists exclusively in the condensed phase. Similar liquid-vapour transitions were also observed in molecular dynamics (MD) simulations of water permeating carbon nanotubes.[42,43]

### 1.2.1 The Hydrophobic Gating Mechanism

This led to the proposal of a hydrophobic gating mechanism,[45] by which the flow of ions through an ion channel is prevented not by steric occlusion of its pore, but instead by the desolvation barrier created by hydrophobic residues along its permeation pathway (Figure 1.4B). In bulk electrolyte, ions are typically associated with a hydration shell comprising several layers of water molecules. Correspondingly, the free energy of hydration is significant for many biological ion species such as $Na^+$ ($-371\,\mathrm{kJ\,mol^{-1}}$),

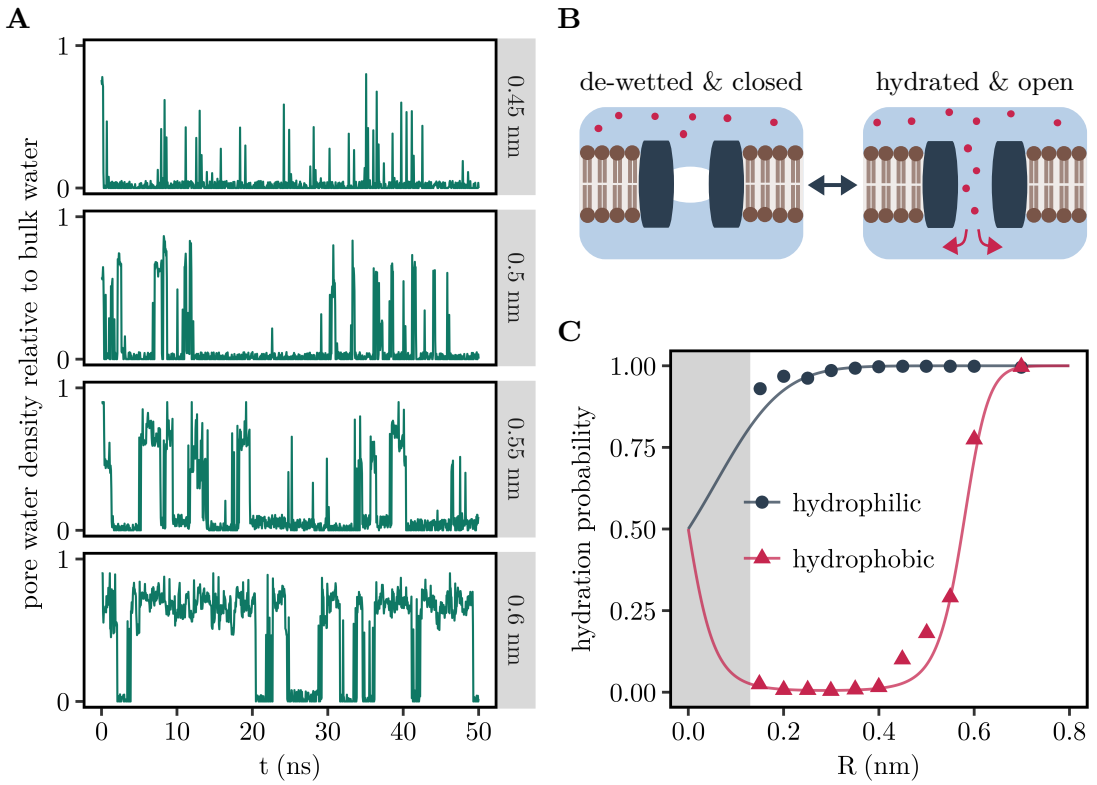



**Figure 1.4:** Liquid-vapour oscillations and the hydrophobic gating mechanism. (A) When confined inside a nanopore with a hydrophobic surface, water density oscillates between a liquid-like and a vapour-like state. The likelihood of the liquid-like state increases with pore radius. (B) Pore de-wetting creates an energetic barrier for ion permeation and functionally closes an ion channel. (C) The probability of the liquid state depends on the radius and hydrophobicity of the pore. A hydrophilic pore remains hydrated so long as its radius exceeds that of a water molecule (∼0.13 nm, indicated by the shaded background), while a hydrophobic pore de-wets at radii below ∼0.6 nm. Panels A and C adapted from Aryal et al. [37] under a creative commons (CC BY 3.0) licence. Original data from Beckstein and Sansom [41] and Beckstein and Sansom [44] respectively. Panel B adapted from Klesse et al. [26] under a creative commons (CC BY 4.0) licence.

$K^+$ ($-298\,\mathrm{kJ\,mol^{-1}}$), or $Cl^-$ ($-373\,\mathrm{kJ\,mol^{-1}}$) and even higher for divalent ions like $Ca^{2+}$ ($-1493\,\mathrm{kJ\,mol^{-1}}$) or $Mg^{2+}$ ($-1820\,\mathrm{kJ\,mol^{-1}}$).[46] Any ion passing a de-wetted channel pore would need to shed its hydration shell and would consequently incur a large energetic cost. Therefore, de-wetting can render an ion channel functionally closed, even if its pore is wide enough to accommodate a hydrated ion without steric clashes.

Further simulation studies have suggested that such a hydrophobic gate can be opened

in several ways.[44] As the hydration state of a pore depends sensitively on both pore radius and surface hydrophobicity (Figure 1.4A), allosteric transitions that either widen an ion channel pore or rotate hydrophobic sidechains away from the pore could induce pore wetting. In fact, the probability of hydrating a pore depends on pore radius in a sigmoidal fashion (Figure 1.4C), which could facilitate a switch-like behaviour where a small increase in radius turns a fully hydrated pore into a fully dehydrated pore. Furthermore, the liquid vapour equilibrium of confined water is also temperature dependent.[44]

Alternatively, a hydrophobic gate may be opened through the application of an external electric field. It was first demonstrated in simulations of model nanopores that the electric field resulting from an ionic concentration imbalance can drive water into an otherwise de-wetted hydrophobic pore, an effect sometimes termed electrowetting.[47,48] Electrowetting has also been observed in simulation studies of carbon nanotubes.[49] Hydrophobic gates may thus also be gated by voltage.

In the preceding review, the concept of hydrophobic gating was introduced on the basis of simulation studies conducted on model nanopore systems. These investigations established that water undergoes liquid-vapour oscillations in hydrophobic confinement and that the de-wetting of a nanopore can create a large energetic barrier for ion permeation. The following section will provide an overview of proposed hydrophobic gates in biological ion channel structures. However, before turning to this survey, a brief note on terminology is in order.

Different authors have used various alternative terms for the mechanism described above, including "hydrophobic gating",[45] "vapour lock",[50] "vapour plug",[51] or "bubble gating".[52] In this thesis, only the expression "hydrophobic gating" is used to denote the mechanism by which ion movement through the de-wetted sections of an ion channel pore is prevented by the free energy barrier associated with stripping the ions of their hydration shell rather than by van der Waals clashes with tightly packed pore residues. It is important to note that this definition of hydrophobic gating excludes cases where pore

de-wetting immediately precedes the collapse of hydrophobic pore segments involved in steric closure due to a conformational change in the protein.[53]

### 1.2.2 Hydrophobic Gates in Biological Ion Channels

Early examples of hydrophobic gating in biological ion channels were found in the mechanosensitive channel of small conductance (MscS), a stress-activated channel that prevents hypo-osmotic shock in bacteria,[6] and in the nicotinic acetylcholine receptor (nAChR), a member of the family of pentameric ligand-gated ion channels (pLGIC) involved in the sensing of neurotransmitters.[54]

The first crystal structure[55] of MscS was initially considered to represent the open state of that channel, as its pore exhibited a minimal radius of >0.3 nm and thus presented no steric impediment to ion conduction. However, two rings of leucine residues (L105 and L109) on the TM3 helices of MscS create a highly hydrophobic environment in the interior of the channel and in MD simulations water was shown to evaporate from its pore on a nanosecond time scale.[56,57] In steered MD simulations, this de-wetting effect resulted in a high free energy barrier for ion permeation, effectively rendering the channel non-conducting.[56] Moreover, MD simulations of an L109S mutant structure, which corresponds to a gain of function phenotype in electrophysiology experiments,[58] resulted in a fully hydrated channel pore.[56]

These observations led to the proposal of a gating mechanism for MscS, in which the rearrangement of the pore-lining helices leads to a dilation of the pore beyond the critical radius required for pore hydration.[56] This prediction was subsequently confirmed by a crystal structure capturing the open conformation of MscS,[59] in which the gate-forming leucine residues are rotated away from the pore, increasing its radius to >0.6 nm and permitting the passage of hydrated ions. Additional simulation studies have since suggested a hydrophobic gate in the related mechanosensitive channel of large conductance (MscL) as well.[60] Moreover, MscS also represents the first example of electrowetting in

a biological ion channel.[61]

As in the case of MscS, the first structure of the closed state nAChR[62] exhibited no physical occlusion of its pore, whose radius is ∼0.3 nm at its narrowest constriction, which is formed by subsequent rings of leucine (L251) and valine (V255) residues. While this is sufficiently large to accommodate a $Na^+$ ion, subsequent MD simulations demonstrated the occurance of liquid-vapour oscillations in this region of the pore.[63] Through detailed free energy calculations it was shown that the hydrophobic L251 and V255 residues act as a hydrophobic gate that creates a large desolvation barrier for permeating ions. This conclusion is supported by numerous electrophysiological studies that have demonstrated an impact of mutations of these residues on nAChR gating.[64–67]

Further support for a hydrophobic gating mechanism in the pLGIC family derives from structures of the prokaryotic ELIC[68] and GLIC[69,70] channels, bacterial homologues of the nAChR. String method MD simulations identified the molecular transition pathway between the open and closed conformations of GLIC and demonstrated that in the former state, water inside the channel pore exists predominantly as a liquid, while in the latter conformation it favours the vapour phase.[71] Furthermore, it was found that the energetic cost of hydrating the pore increases by $18.5\,k_BT$ when transitioning from the open to the closed conformation, compared to an increase of only $2.0\,k_BT$ in the free energy cost for moving a $Na^+$ ion through an already hydrated GLIC pore.[72] The hydration state of the pore is thus the predominant factor in regulating ion flow through GLIC.

With the emergence of an increasing number of novel ion channel structures, the list of putative hydrophobic gates has been expanded beyond these two early examples and now spans across several additional ion channel families. In TWIK-1, a prototypical member of the family of two-pore-domain (K2P) potassium channels, a hydrophobic barrier has been proposed to exist within the inner pore of the channel that restricts permeation.[50] Similarly, de-wetting of a sterically unoccluded pore has been demonstrated to prevent ion flow through the human big potassium (BK) channel.[73] Moreover, the temperature

dependence[44] of the hydration equilibrium in a hydrophobic pore is thought to underlie the temperature sensing ability of the transient receptor potential cation channel TRPV1.[74]

Further examples of hydrophobic gates include the magnesium channel CorA,[75] the calcium release activated calcium channel (CRAC) Orai1,[76] and the calcium activated chloride channel Bestrophin-1.[77] Beyond naturally occuring ion channels, it has also been shown that hydrophobic gates can be engineered into $\beta$-barrel protein nanopores[78] and that such biomimetic porins exhibit electrowetting.[79]

### 1.2.3 Experimental Support for Hydrophobic Gating

Unfortunately, direct observation of water dynamics inside an ion channel pore is not possible with current structural biology techniques. While water molecules are sometimes included in structural models, the presence or absence of water in protein crystals or vitrified protein samples can not provide reliable information about the hydration equilibrium of channel pores at room temperature. Therefore, experimental support for the hydrophobic gating hypothesis comes mainly from electrophysiological mutation studies on ion channels. Additional evidence derives from highly accurate current measurements on artificial nanopores as well as from kinetic models based on double mutant cycle analysis.

As an illustrative example of the first approach, electrophysiological measurements on TWIK-1 were used to validate the existence of a hydrophobic barrier in this potassium-selective background channel.[50] MD simulations had suggested that the inner pore of this channel undergoes de-wetting in a hydrophobic region below the selectivity filter (Figure 1.5A–B) and that this underlies the low rates of ion conduction typically observed for TWIK-1.[80] Through whole cell current recordings, it could be demonstrated that channel activity increases significantly if the hydrophobic leucine residue facing the inner pore of TWIK-1 (L146) is replaced by more hydrophilic asparagine residue (Figure 1.5C).

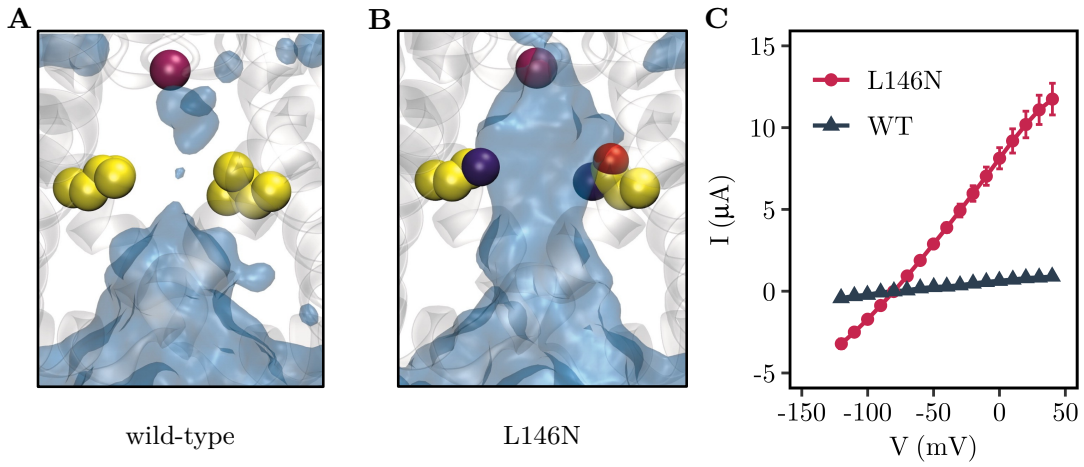



**Figure 1.5:** Hydrophobic gating in the TWIK-1 potassium channel. (A,B) Average water density from MD simulations of (A) wild-type TWIK-1 and (B) a L146N mutant. Substitution of leucine with asparagine leads to hydration of an otherwise de-wetted pore. (C) Current-voltage relationship for whole-cell currents in TWIK-1. While the wild-type channel exhibits low functional activity, the hydrophilic mutation of L146 into asparagine results in significantly larger conductance. Figure adapted from Aryal et al. [37] under a creative commons (CC BY 3.0) licence. Original data from Aryal et al. [50].

Electrophysiological mutation studies have also identified hydrophobic gates in a number of other channels, including MscL[51] and various TRP channels,[81] as well as Piezo1 and Piezo2.[82]

Hydrophobic gating was also observed in artificial nanopores track-etched into polyethylene terephthalate (PET) films, whose inner surface was rendered hydrophobic through the addition of methyl groups by prolonged exposure to trimethylsilyldiazomethane.[83] When these pores were solvated in a KCl electrolyte, ionic currents could only be observed at very high voltages, despite the fact that their radius signficantly exceeded that of either ion species. Importantly, it could be ruled out that the pore was still water-filled in the non-conducting state through current measurements in HCl solution. In a hydrated pore, a proton current would arise via the Grotthus mechanism,[84,85] but could not be measured in this case. It could therefore be concluded that the evaporation of water from the pore renders it impermeable to ions. Furthermore, it was demonstrated that the application of an external electric field promotes the wetting of the PET pores in line

with predictions made by MD simulations of model pores[47,48] and carbon nanotubes.[86]

Double mutant cycle $\phi$-value analysis was used to derive a spatially resolved kinetic model of gating in the nAChR from electrophysiological measurements of channel open probability.[87] In this model, the de-wetting of the gate region of the nAChR pore represents an individual step in the reaction cascade between the closed (C) and open (O) states of the receptor. Specifically, the study described the gating process as a linear chain of four metastable transition states ($T_i$),

$$C \longleftrightarrow T_1 \longleftrightarrow T_2 \longleftrightarrow T_3 \longleftrightarrow T_4 \longleftrightarrow O, \qquad \{1.1\}$$

which are separated by five reversible reactions. While the first three of these steps could be attributed to large scale conformational rearrangements of the protein, the penultimate step was associated with the dilation of the channel gate. The final reaction towards an ion permeable conformation of the channel was then explained as the hydration of the dilated gate. This model represents a refinement of previous kinetic models for nAChR gating[88,89] and provides strong support for a hydrophobic gate in this channel, as was originally proposed by MD simulations of both nAChR[63] and its bacterial homologue GLIC.[71,72]

### 1.2.4 Thermodynamic Theory of Hydrophobic Gating

From a thermodynamic perspective, the stochastic oscillations between the liquid and vapour states of confined water can be understood through the theory of capillary evaporation and condensation.[41] A necessary condition for the existence of water vapour inside a confining cavity is that the water pressure, $P$, exceeds the saturation vapour pressure, $P_{\mathrm{sat}}$, of the bulk liquid. Both pressures are related through the Kelvin equation,

$$\ln \frac{P}{P_{\mathrm{sat}}} = -\beta \gamma_{\mathrm{lv}} \frac{V_{\mathrm{mol}}}{r}, \qquad (1.1)$$

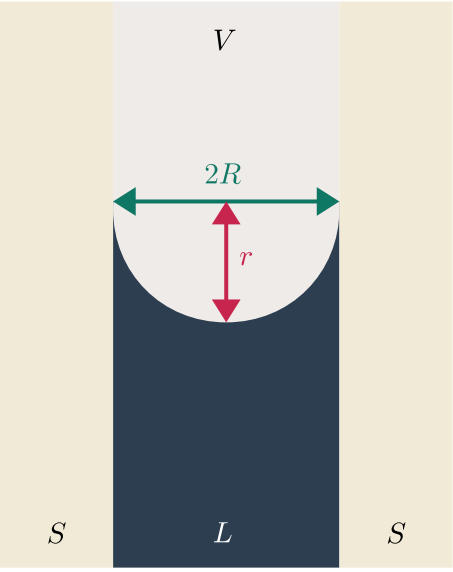



**Figure 1.6:** Liquid-vapour coexistence in cylindrical capillaries. The interface between the liquid ($L$) and vapour ($V$) phases of water assumes a meniscus shape with curvature radius $r$ when confined in a cylindrical pore of radius $R$. When the confining solid surface ($S$) is hydrophobic, the meniscus is concave, as shown here, otherwise it is convex. In thermodynamic equilibrium, both vapour and liquid have the same pressure, i.e. $P_\mathrm{v} = P_\mathrm{l} = P$. If this pressure exceeds the saturation vapour pressure, liquid water and water vapour can coexist inside the pore, otherwise it must be filled with liquid.

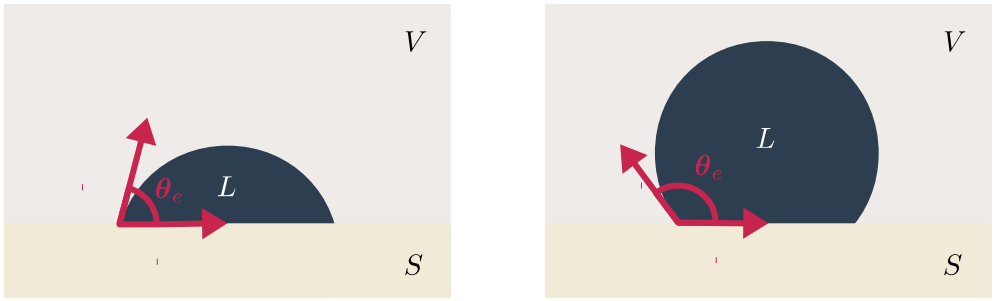
where $\beta^{-1} = k_\mathrm{B}T$ is the temperature factor, $\gamma_\mathrm{lv}$ is the surface tension at the liquid-vapour interface, $V_\mathrm{mol}$ is the molecular volume of water, and $r$ is the radius of curvature of the interfacial meniscus (Figure 1.6).

In a cylindrical pore, the Kelvin equation can be re-expressed as

$$\ln \frac{P}{P_\mathrm{sat}} = -\beta(\gamma_\mathrm{sv} - \gamma_\mathrm{sl})\frac{V_\mathrm{mol}}{R}, \tag{1.2}$$

where $\gamma_\mathrm{sv}$ and $\gamma_\mathrm{sl}$ denote the surface tensions at the solid-vapour and solid-liquid interfaces respectively, and $R$ is the radius of the pore. From Equation 1.2 it can be seen that the evaporation criterion ($P > P_\mathrm{sat}$) will be met only if the formation of a solid-vapour interface is energetically favourable compared to the formation of a solid-liquid interface ($\gamma_\mathrm{sl} > \gamma_\mathrm{sv}$). In other words, water vapour can exist in a pore, if its surface is hydrophobic.

This insight naturally leads to the development of a thermodynamic model for the probability of pore hydration.[40,44] For a classical two-state system obeying Boltzmann




hydrophilic surface hydrophobic surface

**Figure 1.7:** Definition of contact angle. Droplets of liquid water ($L$) on solid surfaces ($S$) and in equilibrium with water vapour ($V$) assume a shape of minimal total surface energy. On a hydrophilic surface, this leads to an acute contact angle, $\theta_c < 90°$. On a hydrophobic surface, water minimises the area of the liquid-surface interface, leading to an obtuse contact angle, $\theta_c > 90°$.

statistics, the probability that a nanopore exists in a liquid state is given by,

$$P_{\text{hydr}} = \frac{1}{1 + \exp\left(-\beta \Delta\Omega_{\text{lv}}\right)} \tag{1.3}$$

where $\Delta\Omega_{\text{lv}} = \Omega_{\text{v}} - \Omega_{\text{l}}$ is the difference between the free energy of the liquid state, $\Omega_{\text{l}}$, and the vapour state, $\Omega_{\text{v}}$. Here $\Omega$ denotes the grand potential of the system, as the number of water molecules inside the pore can vary through particle exchange with the bulk domain on either side. When $\Delta\Omega_{\text{lv}} < 0$, the pore will be predominantly filled with vapour, while for $\Delta\Omega_{\text{lv}} > 0$, the liquid phase is favoured.

It can be shown that for short nanopores (length $L < 10\,\text{nm}$) and ambient pressure, the free energy difference is dominated by surface terms.[44] For a cylindrical pore, it is then given by

$$\Delta\Omega_{\text{lv}} = 2\pi\gamma_{\text{lv}} R(R + L\cos\theta_{\text{c}}), \tag{1.4}$$

where $\theta_{\text{c}}$ denotes the contact angle of a water droplet adsorbed on a surface (Figure 1.7). The free energy difference can thus only become negative in hydrophobic pores,

where $\theta_c > 90°$. Moreover, Equation 1.4 highlights that the maximum radius for which de-wetting can occur increases with the length of the pore.

It needs to be pointed out that the preceding treatment is phenomenological, since the Kelvin and Young equations are only strictly valid at macroscopic length scales. However, it has been demonstrated that Equation 1.4 accurately describes the probability of pore hydration in MD simulations of cylindrical pore models with radii in the sub-nanometer range (Figure 1.4B). Therefore, even though the pores of biological ion channels are not ideal cylinders, this model nonetheless provides a powerful framework for understanding hydrophobic gating.

In addition, Equation 1.4 can be extended to describe the electrowetting of nanopores as well.[47] Since the capacitance of a volume element is proportional to the relative permittivity of the material filling it, it is possible to store more electrical energy in a liquid-filled pore than in a de-wetted cavity. Under the influence of an external electric field, $E$, the free energy between the liquid and vapour states of a nanopore therefore becomes

$$\Delta\Omega_{lv} = 2\pi\gamma_{lv}R(R + L\cos\theta_c) + \frac{1}{2}\varepsilon_0(\varepsilon_l - \varepsilon_v)VE^2, \tag{1.5}$$

where $V = \pi LR^2$ denotes the pore volume, $\varepsilon_v \approx 1$ and $\varepsilon_l \approx 80$ represent the relative permittivities of water vapour and liquid water respectively, and $\varepsilon_0$ is the dielectric constant. As the contribution of the electric energy is positive, it is always possible to hydrate an otherwise de-wetted pore, provided that the electric field is sufficiently strong.

## 1.3 Molecular Dynamics for Functional Annotation

As has been demonstrated in the previous section, hydrophobic gating is clearly an important process for ion channels. However, it also constitutes a serious challenge for

structural biologists seeking to infer the molecular mechanisms underlying ion channel function from structural models. In particular, the potential existence of a hydrophobic gate complicates the assessment of whether a given structure corresponds to an open (ion conducting) or closed (non-conducting) conformation of the channel and thus hampers the comparison of structural and electrophysiological data.

Currently, the functional status of an ion channel structure is assigned based on the presence or absence of constrictions along its permeation pathway that would prohibit the passage of an ion through steric clashes. However, the physical dimensions of the pore can not be directly mapped to the functional state of a channel, as a hydrophobic pore with an opening wide enough for an ion to pass through may nonetheless be functionally closed. In light of the rapidly growing number of experimentally determined channel structures (Figure 1.3), there is thus a clear need for a novel method that takes into account the de-wetting behaviour of hydrophobically confined water.

Computational approaches provide an opportunity to bridge the gap between structural and functional data, and atomistic MD simulations in particular have provided valuable insights into structure-function relationships in ion channels.[32,35,90] Several methods for investigating conduction in ion channels have been proposed, including the calculation of ionic free energy profiles via umbrella sampling[91] and the direct measurement of ionic currents under the influence of an external electric field.[92] More recently, the latter technique has evolved into the computational electrophysiology method,[93] where a transmembrane voltage arises naturally from an ionic concentration gradient, representative of physiological conditions.

Unfortunately, despite continuing advancements in high performance computing hardware and simulation software,[94–97] atomistic MD simulations of membrane proteins are still computationally expensive and complex techniques such as computational electrophysiology are thus not yet well suited for the high-throughput analysis of large numbers of ion channel structures. However, it has been shown that a reliable assessment of the

functional status of a channel structure can be achieved through simpler equilibrium MD simulations, by using the behaviour of water in the channel pore as a proxy for ionic permeability.[78]

A central aim of this thesis therefore was the development of a robust and versatile software tool for the functional annotation of ion channel structures. This programme, now named Channel Annotation Package (CHAP), provides an easy-to-use implementation of the simulation protocol developed by Trick et al. [78] and was recently made available to the scientific community.[26] CHAP combines the ability of existing tools to measure the dimensions of ion channel pores with a simulation-based analysis of the behaviour of pore water by leveraging the MD engine GROMACS.[98] It thus provides a computationally efficient means of screening novel ion channel structures for both hydrophobic and conventional steric gates and enables the rapid functional annotation of ion channel structures.

As CHAP is based on MD simulations and thus relies on empirical force fields, it is also important to ascertain the validity of these force fields for the typical conditions in an ion channel pore. As has been pointed out by several authors,[60,63,72] the main limitation of the MD simulations of hydrophobic gates carried out thus far is that they did not explicitly account for the effect of electronic polarisation, relying instead on a mean field treatment. This issue is addressed in this thesis and the influence of electronic polarisation is quantified through MD simulations employing advanced polarisable force fields.[99,100]

## 1.4 Thesis Outline

This thesis documents the development of a new computational tool for the functional annotation of ion channels and its application to several recently determined structures of two neurotransmitter receptors. It also presents one of the first efforts to explicitly include the effect of electronic polarisability in a molecular dynamics investigation of

biological ion channels*. The presentation of these results is structured as follows.

In Chapter 2, an overview of the theory underlying the computational methods employed throughout this thesis is presented. Here the focus lies on atomistically resolved molecular dynamics simulations including a survey of next-generation force fields with the ability to explicitly model induced polarisation. An introduction to continuum electrostatics calculations within the Poisson-Boltzmann framework is also provided.

Based on work published in Klesse et al. [26], Chapter 3 introduces the Channel Annotation Package, CHAP. The chapter begins with an overview of how CHAP can be used either as a standalone tool to detect hydrophobic pore constrictions in experimentally determined ion channel structures or in conjunction with MD simulations to identify hydrophobic gates by analysing the dynamical behaviour of water. This is followed by a description of the implementation of CHAP as a C++ programme, including a discussion of the numerical algorithms used for the detection of permeation pathways, the estimation of local water densities, and the calculation of hydrophobicity profiles.

Subsequently, the utility of CHAP is demonstrated in Chapter 4, which describes the identification of hydrophobic gates in two members of the family of pentameric ligand-gated ion channels, the 5-$HT_3$ and glycine receptors. This chapter is partly based on results published by Trick et al. [78], but is extended to include additional structures of these channels not yet available in 2016. It also employs a more recent version of CHAP than the early prototype used by Trick et al. [78]. In addition to discussing two particular channels in detail, Chapter 4 also provides a glimpse of the larger picture and discusses the prevalance of hydrophobic gates in the ion channel structural proteome based on results published in Rao et al. [103]. This study employed CHAP for the high throughput screening of nearly 200 ion channel structures and elucidates under what conditions of pore size and surface character hydrophobic gating arises in biological ion

*The only previous study to employ a polarisable force field concerns the gramicidin A channel.[101] It should, however, be noted that the influence of electronic polarisation on ion channel function has previously been explored through a QM/MM approach.[102]

channels.

Chapter 5 explores the influence of electronic polarisation in MD simulations of ion channels alongside a systematic comparison of modern water models. It is found that inside a hydrophobic gate, water experiences a significant reduction of its molecular dipole moment, comparable in magnitude to the difference between liquid water and water vapour. Ions also behave differently when electronic polarisation is explicitly modelled and generally interact more strongly with the pore-forming protein residues, an effect that is particularly pronounced for $Cl^-$. These results emphasise the importance of physically accurate models in quantitative investigations of ion channel function.

The electrowetting effect is examined in Chapter 6. Here it is demonstrated that pore de-wetting prevents ion permeation through the 5-$HT_3$ receptor and that a closed hydrophobic gate can only be hydrated by a superphysiological electric field well in excess of that which an ion channel would normally experience. Furthermore, the kinetic rate constants for transitions between the liquid and vapour states of a channel pore are shown to be inversely correlated when the free energy of pore hydration is varied, eliminating the concern that liquid-vapour oscillations may occur on very long time scales inaccessible by equilibrium MD simulations.

Finally, the main conclusions of this thesis are summarised in Chapter 7 together with a discussion of future research directions.

I like this place, and willingly
could waste my time in it.

*(William Shakespeare)*

# 2

# Theory and Methods

The purpose of this chapter is to provide the theoretical and computational background of the biomolecular simulation methods underlying the results presented in this thesis. Classical molecular dynamics (MD) is first introduced in the context of conventional atomistic simulations with pairwise additive force fields, before discussing the inclusion of induced polarisation into this framework. This is followed by a brief description of how thermodynamic free energies can be derived from the results of an MD simulation. An overview of the Poisson-Boltzmann model for continuum electrostatics calculations concludes this chapter.

## 2.1 Classical Molecular Dynamics

Classical MD is a computer simulation technique for determining the time evolution of a set of atomic positions, $\{\mathbf{r}_i\}$, and momenta, $\{m_i\dot{\mathbf{r}}_i\}$, under the influence of a conservative potential, $U(\{\mathbf{r}_i\})$. This is accomplished by numerically solving the classical equation of motion,

$$\ddot{\mathbf{r}}_i(t) = -\frac{1}{m_i}\nabla_{\mathbf{r}_i}U(\{\mathbf{r}_i(t)\}), \tag{2.1}$$

where $m_i$ denotes particle mass and the index $i$ is used — as in the remainder of this chapter — to refer to the $i$-th particle. The resulting set of time-dependent particle positions and velocities is referred to as a trajectory and provides a sample of microstates of the simulated system. From this sample, the relevant thermodynamic properties of a biomolecular system, such as the free energy difference between pore hydration states, can be derived through simple time averaging. Provided that care is taken in the treatment of temperature and pressure, kinetic and dynamical properties of the system, such as the conductance of an ion channel, can be extracted from an MD trajectory as well.

The potential $U(\{\mathbf{r}_i\})$ in Equation 2.1 provides a classical approximation to the quantum-mechanical interaction energy of a system of simulated atoms. Its specific functional form is chosen empirically to capture the relevant quantum effects. The associated parameters for specific atom types are then determined from an iterative fitting procedure to both experimental data and ab initio quantum mechanical calculations. The combination of a functional form and an optimised parameter set is known as a *force field*.

Due to their empirical nature, force field parameters need to be optimised individually for each molecular species. For simulations of chemically heterogeneous systems, a consistent force field parameterisation for all relevant species is required. This has led to the development of several consistently parameterised force fields for biomolecular simulation, which provide parameters for the most relevant biochemical species, typically including amino acids, nucleic acids, lipids, carbohydrates, ions, and water. These biomolecular force fields differ in how they balance physical accuracy with computational efficiency and can be divided into two categories based on whether many-body effects are included in their description of intermolecular interactions.

The remainder of this section will focus on the first category comprising the *pairwise additive* force fields. In these, the energy due to the interaction of two non-bonded atoms

is independent of the position of all other atoms. By contrast, the second category permits general $N$-body interactions, which allows the inclusion of a greater variety of physical effects, but is computationally significantly more costly. *Polarisable* force fields, which model induced atomic dipoles explicitly, belong to this category and will be discussed in Section 2.2.

### 2.1.1 Pairwise Additive Force Fields

Pairwise additive force fields can be further subdivided into force fields that model molecules in full atomistic detail and united-atom force fields that reduce the effective number of degrees of freedom by merging hydrogen atoms with the heavy atoms to which they are bonded. The former subgroup includes the Assisted Model Building and Energy Refinement (AMBER)[104] and Chemistry at Harvard Macromolecular Mechanics (CHARMM)[105] force fields, while notable examples of the latter are the Groningen Molecular Simulation (GROMOS)[106] and Optimized Potential for Liquid Simulations united-atom (OPLS-UA)[107] force fields. Coarse-grained force fields like MARTINI,[108] which represent collections of multiple heavy atoms and their bonded hydrogens as a single particle, also exist.

The majority of simulations carried out as part of this work employ the latest version of the CHARMM family of force fields: CHARMM36m.[109] This force field was chosen because it provides parameters for all required molecules (including a wide variety of lipids) and has been demonstrated to be well suited for the simulation of membrane proteins.[110] In addition, the CHARMM Drude polarisable force field employs a functional form similar to that used by CHARMM36m, making this force field a sensible reference case against which results from polarisable simulations can be compared. The types of intermolecular (non-bonded) and intramolecular (bonded) interactions included in the CHARMM36m force field are illustrated in Figures 2.1 and 2.2 and their functional form is discussed below.

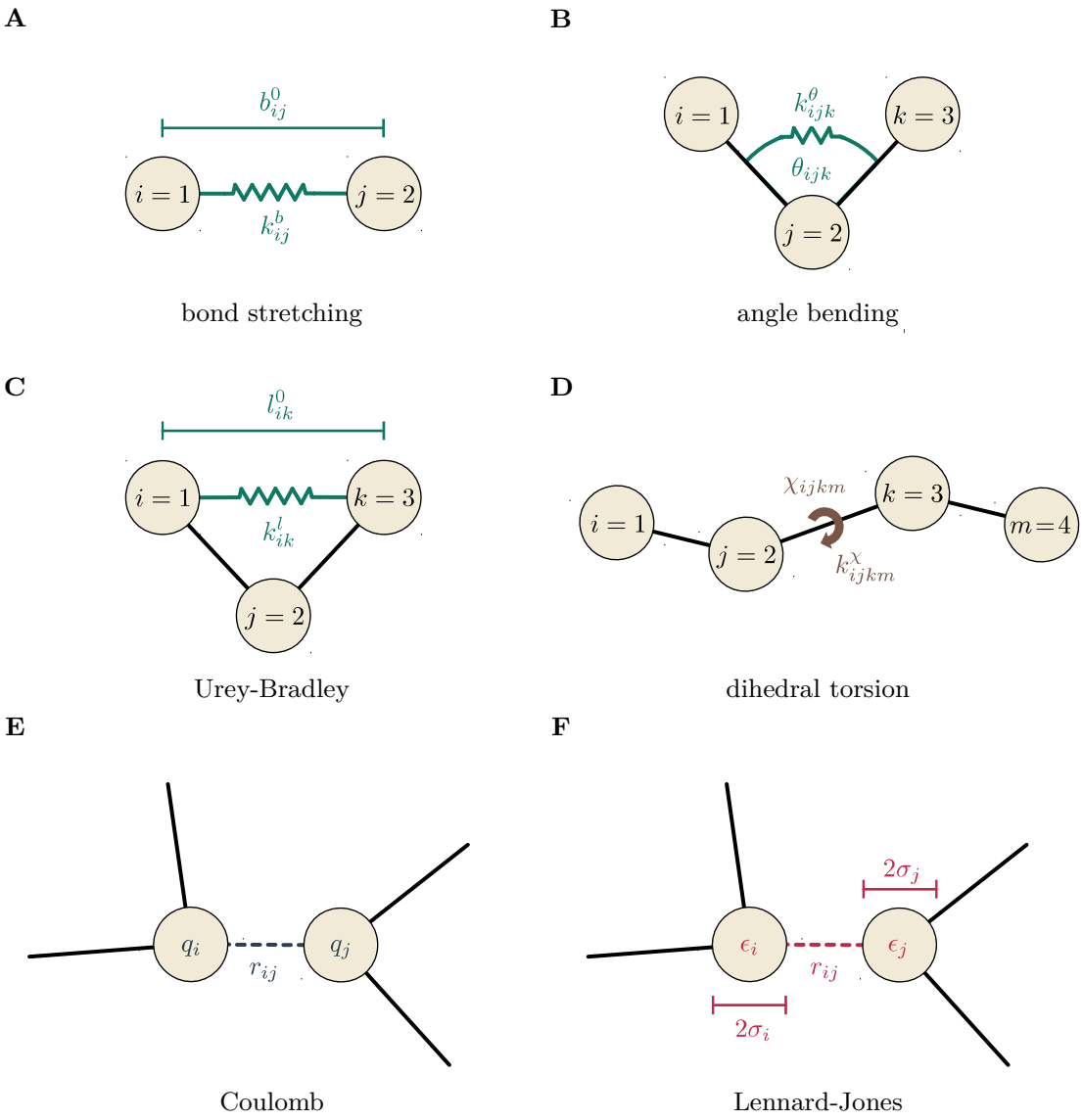



**Figure 2.1:** Fundamental interaction types of common molecular dynamics force fields. (A,B) Bond stretching and angle bending are typically modelled as harmonic oscillators, although some force fields include anharmonic terms. (C) A Urey-Bradley term is a harmonic potential applied to the distance between a 1–3 pair of atoms. (D) Rotations around bonds are subjected to a periodic dihedral torsion potential. (E) Partial charges interact via the electrostatic Coulomb potential. (F) Van der Waals forces are modelled through the Lennard-Jones potential. In the CHARMM36m force field, Coulomb and Lennard-Jones interactions are not calculated for 1–2 and 1–3 pairs of atoms and may be scaled for 1–4 pairs of atoms.

### Intermolecular Interactions

In the (pairwise additive) CHARMM36m force field, non-bonded interactions between two particles comprise the electrostatic Coulomb potential (Figure 2.2C), $U_{\text{ele}}$, and a van der Waals term, $U_{\text{vdW}}$. The Coulomb potential has the form

$$U_{\text{ele}} = \frac{1}{4\pi\epsilon_0}\frac{q_i q_j}{r_{ij}}, \tag{2.2}$$

where $\epsilon_0$ is the permittivity of vacuum, $q_i$ is the charge of the $i$-th atom, and $r_{ij}$ is the distance between atoms $i$ and $j$. Note that most atoms carry a non-integer partial charge, reflecting the fact that electrons shared between covalently bonded atoms are shifted towards the more electronegative partner.

Van der Waals forces are modelled through the Lennard-Jones potential (Figure 2.2D),

$$U_{\text{vdW}} = 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right], \tag{2.3}$$

which is strongly repulsive for short interatomic separations and weakly attractive beyond a critical distance $R_{ij}^{\text{vdW}} = 2^{1/6}\sigma_{ij}$. The coefficients $\epsilon_{ij}$ and $\sigma_{ij}$ control the depth of the potential well and the distance at which the potential is zero respectively. Their values are calculated using the Lorenz-Berthelot combining rules,

$$\epsilon_{ij} = \sqrt{\epsilon_i \epsilon_j} \tag{2.4}$$

and

$$\sigma_{ij} = \frac{1}{2}(\sigma_i + \sigma_j) \tag{2.5}$$

where $\epsilon_i$ and $\sigma_i$ are per-atom parameters of the force field. The repulsive part of the Lennard-Jones potential accounts for the quantum-mechanical exchange interaction that

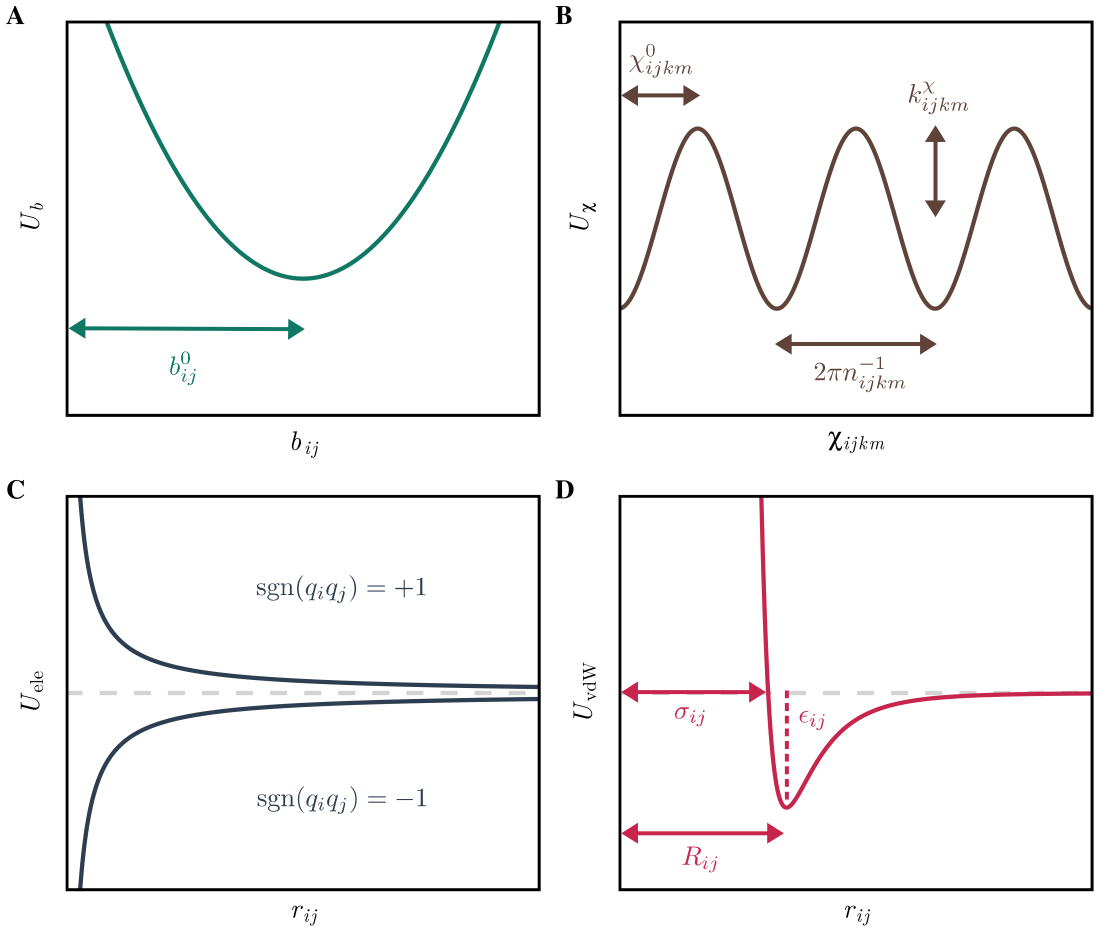



**Figure 2.2:** Functional forms of common molecular dynamics force field terms. (A) Harmonic bond stretching potential. Angle bending and Urey-Bradley forces use the same functional form. (B) Periodic dihedral torsion potential. (C) Coulomb potential for a pair of charges. If both charges have the same sign, the electrostatic interaction is repulsive (top), otherwise it is attractive (bottom). (D) At short distances, the van der Waals potential is strongly repulsive. At larger distances, dispersion interactions dominate and the potential is weakly attractive.

arises when the electronic clouds of adjacent atoms overlap, while its attractive part represents the dispersion interaction due to mutually induced dipoles.

### Intramolecular Interactions

Intramolecular forces control the stretching and bending of covalent bonds as well as torsional rotations in a chain of consecutively bonded atoms. They are modelled through

a set of specialised $N$-body potentials between neighbouring atoms. In this context, two directly bonded atoms are referred to as a 1–2 pair, while atoms separated by two and three bonds are referred to as 1–3 pair and 1–4 pair respectively. In the CHARMM force field, 1–2 and 1–3 pairs are excluded from the calculation of non-bonded forces, as their interaction is assumed to be fully prescribed by the intramolecular potential functions discussed below. By contrast, 1–4 pairs do contribute to the non-bonded potential, but van der Waals forces between specific 1–4 pairs may be scaled.

The elongation of covalent bonds is treated through a harmonic potential (Figure 2.2A),

$$U_b = k_{ij}^b (r_{ij} - b_{ij}^0)^2, \tag{2.6}$$

where $k_{ij}^b$ is a force constant and $b_{ij}^0$ is the reference bond length. An identical functional form is used for the Urey-Bradley potential,

$$U_l = k_{ik}^l (r_{ik} - l_{ik}^0)^2, \tag{2.7}$$

which acts between 1–3 pairs of atoms, where $k_{ik}^l$ and $l_{ik}^0$ denote the associated force constant and reference distance. The bond angle

$$\theta_{ijk} = \arccos\left(\frac{\mathbf{r}_{ij} \cdot \mathbf{r}_{jk}}{r_{ij} r_{jk}}\right) \tag{2.8}$$

defined by a 1–3 pair of atoms ($i$ and $k$) and their mutual partner ($j$) is also subjected to a harmonic potential,

$$U_\theta = k_{ijk}^\theta (\theta_{ijk} - \theta_{ijk}^0)^2, \tag{2.9}$$

with force constant $k_{ijk}^\theta$ and reference angle $\theta_{ijk}^0$.

In practice, these interactions are significantly stronger than Coulomb or van der

Waals forces. Consequently, bond lengths and angles as well as 1–3 pair distances never deviate substantially from their respective reference value. This in turn justifies the use of harmonic potentials, which can be interpreted as the leading term in a Taylor series approximation of the true potential around the reference distance or angle.

Torsional motions, by contrast, may be large. They are modelled through a periodic dihedral potential (Figure 2.2B),

$$U_\chi = k^\chi_{ijkm}\left(1 + \cos(n_{ijkm}\chi_{ijkm} - \chi^0_{ijkm})\right), \tag{2.10}$$

where $k^\chi_{ijkm}$ is a force constant, $n_{ijkm}$ is the number of minima along a full rotation around the bond between atoms $j$ and $k$, and $\chi^0_{ijkm}$ is a phase shift. By convention, the dihedral angle is defined as

$$\chi_{ijkm} = \arccos\left(\frac{\mathbf{r}_{ji} \times \mathbf{r}_{jk}}{||\mathbf{r}_{ji} \times \mathbf{r}_{jk}||} \cdot \frac{\mathbf{r}_{km} \times \mathbf{r}_{jk}}{||\mathbf{r}_{km} \times \mathbf{r}_{jk}||}\right) \tag{2.11}$$

in terms of the positions of a linear chain of four atoms ($i$, $j$, $k$, and $m$).

An improper dihedral angle, $\phi_{ijkm}$, can be defined analogously, but in this case the underlying atoms do not need to be sequentially bonded to one another. Improper dihedrals are subjected to a harmonic potential,

$$U_\phi = k^\phi_{ijkm}(\phi_{ijkm} - \phi^0_{ijkm})^2, \tag{2.12}$$

with force constant $k^\phi_{ijkm}$ and reference angle $\phi^0_{ijkm}$ in order to enforce the planarity of aromatic rings.

### 2.1.2 Water Models

An important consideration in biomolecular simulation is the choice of water model. One common approach is to account for the influence of solvation implicitly by modelling

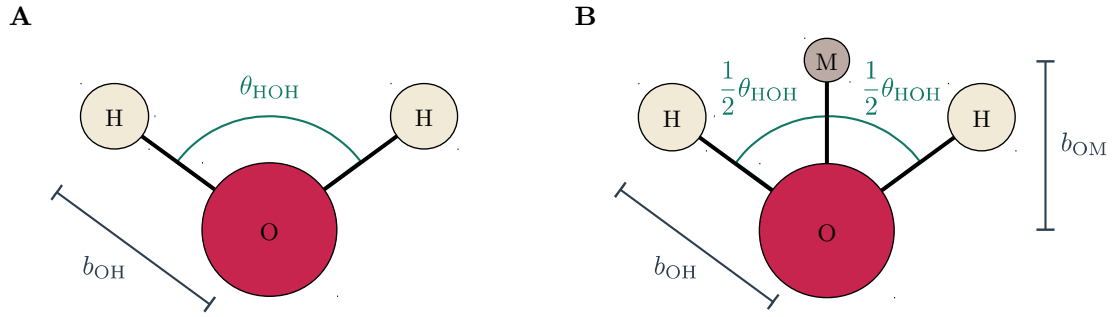



**Figure 2.3:** Geometry of common explicit water models. (A) Three site water models consist of three Coulomb interaction sites representing the hydrogen (white) and oxygen (red) atoms of the water molecule with charges $q_H$ and $q_O = -2q_H$ respectively. There is typically only one van der Waals interaction site at the position of the oxygen ($\epsilon_O$, $\sigma_O$), but some models also include van der Waals sites for the hydrogen atoms ($\epsilon_H$, $\sigma_H$). (B) Four site water models comprise an additional virtual site (light brown) located on the H–O–H bisector. This "M"-site then carries a negative charge ($q_M$), while the oxygen atom interacts only via van der Waals forces. Sizes and distances are not drawn to scale, see Table 2.1 for the geometric parameters of additive water models.

water as a structureless dielectric continuum either through the semi-analytical generalised Born framework[111] or by numerically solving the Poisson-Boltzmann equation.[112] However, in the present context of studying the behaviour of water in hydrophobic confinement, where entropic effects are important, an explicit water model is required.

Explicit water models typically consist of three or four interaction sites for Coulomb and van der Waals forces, although models with five[113] or six[114] sites also exist. Their geometry reflects the chevron shape of the water molecule (Figure 2.3) and is typically held rigid during MD simulations in order to suppress high frequency oscillations of water hydrogens. Within this rigid point charge framework, a large number* of water models has been developed and a comprehensive review is beyond the scope of this chapter. Instead, a brief overview of the most relevant models for the purpose of biomolecular simulation is given and the reader is referred to a recent review[117] for a more in-depth discussion.

Biomolecular water models are typically targeted at reproducing the behaviour of

*A review[115] from the year 2002 lists 32 empirical water models. Their number has grown significantly over the last two decades and SklogWiki,[116] an online resource, lists 111 water models as of 2019.

| Parameter | Units | SPC/E[122] | mTIP3P[120] | TIP4P[119] | TIP4P/2005[123] | OPC[124] |
|---|---|---|---|---|---|---|
| $N_{\text{sites}}$ | — | 3 | 3 | 4 | 4 | 4 |
| $b_{\text{OH}}$ | nm | 0.1000 | 0.0957 | 0.0957 | 0.0957 | 0.0872 |
| $b_{\text{OM}}$ | nm | — | — | 0.0150 | 0.0155 | 0.0159 |
| $\theta_{\text{HOH}}$ | ° | 109.47 | 104.52 | 104.52 | 104.52 | 103.60 |
| $q_{\text{O}}$ | $e$ | −0.8476 | −0.8340 | — | — | — |
| $q_{\text{H}}$ | $e$ | 0.4238 | 0.4170 | 0.5200 | 0.5564 | 0.6791 |
| $q_{\text{M}}$ | $e$ | — | — | −1.0400 | −1.1128 | −1.3582 |
| $\sigma_{\text{O}}$ | nm | 0.3166 | 0.3151 | 0.3154 | 0.3159 | 0.3167 |
| $\sigma_{\text{H}}$ | nm | — | 0.0400 | — | — | — |
| $\epsilon_{\text{O}}$ | $\text{kJ mol}^{-1}$ | 0.6500 | 0.6364 | 0.6480 | 0.7749 | 0.8903 |
| $\epsilon_{\text{H}}$ | $\text{kJ mol}^{-1}$ | — | 0.1925 | — | — | — |

**Table 2.1:** Geometry, electrostatic, and van der Waals parameters of the additive water models employed in this thesis. Note that the original TIP3P water model[118] uses the same parameters as the CHARMM-modified mTIP3P water model, except that it does not have van der Waals interaction sites on the hydrogen atoms and hence no $\sigma_{\text{H}}$ or $\epsilon_{\text{H}}$. See Figure 2.3 for the definition of these parameters.

liquid water at ambient conditions. The most important amongst these are the models that were used in the parameterisation of the major biomolecular force fields discussed in the previous section. The AMBER, CHARMM, and OPLS force fields are all based on the transferable intermolecular potential (TIP) family of models. The three site model TIP3P[118] was employed in the development of the AMBER force fields, while the four site model TIP4P[119] was used to optimise the parameters of the OPLS force field. CHARMM is based on a modified version of the TIP3P model, referred to as mTIP3P,[120] which carries additional van der Waals interaction sites on the hydrogen atoms. The GROMOS force fields on the other hand are based on the simple point charge (SPC)[121] water model, which like TIP3P comprises three interaction sites, but with a wider HOH angle based on the geometry of ice.

A plethora of refined water models has been developed on the basis of these host water models. A considerable improvement of the SPC model was achieved through

the inclusion of a self-energy term, yielding the SPC/E[122] model. Reparameterisations of the TIP family aimed for compatibility with Ewald electrostatics[125,126] as well as the development of flexible variants.[127,128] Another notable reparameterisation is the TIP4P/2005[123] general purpose model, which has been found to be particularly accurate in a number of benchmark studies.[129–131] More recently, the optimal point charge (OPC) model family was introduced. It comprises the four site OPC[124] model and the three site OPC3[132] model, which both show improved accuracy compared to other rigid point charge water models.

Most simulations in this thesis were carried out using the CHARMM36m force field with the mTIP3P water model. However, the sensitivity of the results to the choice of water model was investigated using a number of alternative water models. These include the traditional TIP4P and SPC/E models as well as the modern TIP4P/2005 and OPC models. The geometry and interaction parameters of these models are summarised in Table 2.1.

### 2.1.3 Molecular Dynamics Algorithms

Having specified the potential $U(\{\mathbf{r}_i\})$, through which individual atoms interact, above, this section describes how Equation 2.1 is solved numerically during an MD simulation. The workflow of such a simulation is summarised in Figure 2.4 and the relevant algorithms are described below.

#### Time Integration

A common discretisation of the equation of motion is the leap frog integrator,[133] in which the position, $\mathbf{r}_i$, and velocity, $\dot{\mathbf{r}}_i$, of a particle are evolved over time according to

$$\dot{\mathbf{r}}_i\left(t + \Delta t/2\right) = \dot{\mathbf{r}}_i\left(t - \Delta t/2\right) + \Delta t\, \ddot{\mathbf{r}}_i(t) \tag{2.13}$$

$$\mathbf{r}_i(t + \Delta t) = \mathbf{r}_i(t) + \Delta t\, \dot{\mathbf{r}}_i\left(t + \Delta t/2\right) \tag{2.14}$$

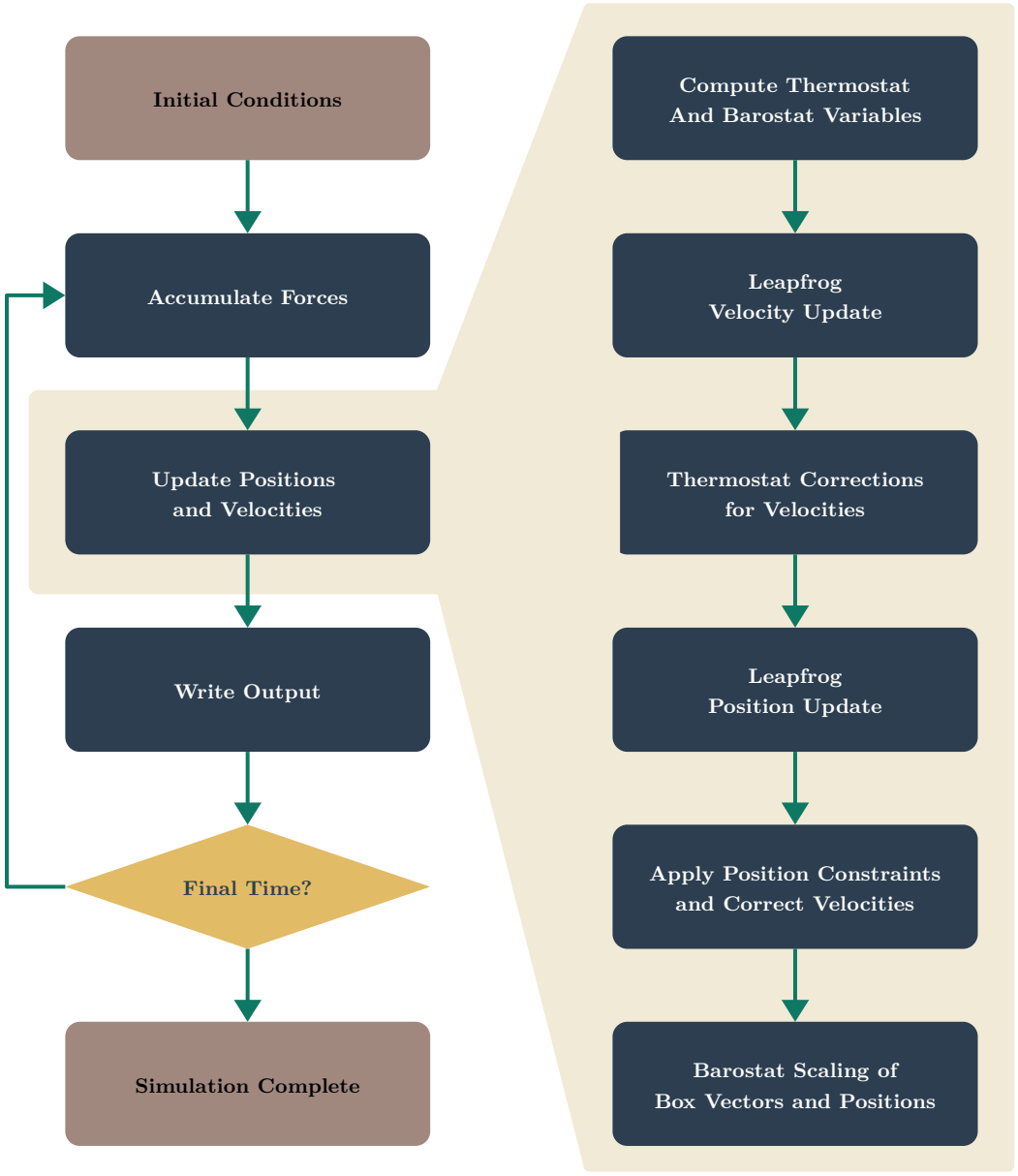



**Figure 2.4:** Workflow of a typical MD simulation. Starting from a set of initial positions and velocities, a trajectory is generated through an iterative three step procedure. In a first step, the forces acting on each particle are accumulated by evaluating the force field potential. Subsequently, the corresponding accelerations are used to update particle positions and velocities by numerically integrating the equation of motion. Finally, all quantities of interest are written to an output file. This procedure is repeated until a specified final time is reached. In practice, the position and velocity update involves additional steps to ensure that the simulation is carried out at the desired temperature and pressure and that constraints on molecular geometry are enforced. However, the most computationally demanding step is the calculation of forces and in particular the evaluation of long-ranged intermolecular interactions.

where $\Delta t$ denotes the time step, and $\ddot{\mathbf{r}}_i(t)$ is the acceleration arising from the potential gradient at the current position of the particle (Equation 2.1). The leapfrog integrator is time-reversible and symplectic, i.e. it preserves the total energy of the system. This property, combined with its relative accuracy and simplicity,[134] have made the leapfrog algorithm the de-facto standard algorithm for MD simulations and it is available in most software packages including GROMACS and OpenMM.

For stable integration with the leapfrog method, the time step must be small enough to capture the fastest dynamical process in the simulation system. This is usually the vibration of covalent bonds, especially those involving lightweight hydrogen atoms, which restricts the time step to at most 1 fs. However, bond vibrations themselves are typically of little interest in the context of biomolecular simulation. It is therefore common practice to constrain these bonds to their equilibrium length, which enables a larger time step of 2 fs. Constraints can also be applied to bond angles, but this may cause artefacts in conformational dynamics,[135] and angle constraints are not employed in this thesis.

A variety of algorithms for applying such holonomic constraints while still representing particle positions in Cartesian coordinates has been developed. The most common procedure is the SHAKE algorithm,[136] which incorporates constraints into the equation of motion through Lagrange multipliers and solves the resulting nonlinear constraint equations iteratively. For the special case of water molecules, the Lagrangian constraint equations can be solved analytically, resulting in the SETTLE method.[137] A faster general alternative is the non-iterative LINCS[138,139] algorithm, which enforces constraints through a two step procedure of first updating bond lengths and subsequently correcting for bond rotations.

#### Temperature and Pressure Control

Integration of the equations of motion as described above leads to a simulation in which particle number, internal energy, and volume are constant. Correspondingly, conforma-

tions sampled during such a simulation are representative of the microcanonical (NVE) ensemble. However, biomolecular processes of interest typically occur under conditions of constant temperature, $T$, and pressure, $P$, and simulations should therefore sample from the isothermal-isobaric (NPT) ensemble.

The Andersen thermostat[140] enables temperature control by randomising particle velocities in accordance with the Maxwell-Boltzmann distribution. In each time step, the $i$-th particle is selected for velocity randomisation with probability

$$P_i = 1 - \exp\left(-\nu \Delta t\right), \tag{2.15}$$

where $\nu$ can be interpreted as the frequency of collisions with particles in a heat bath. The particle velocity is then set to

$$\dot{\mathbf{r}}_i = \sqrt{\frac{k_\mathrm{B} T}{m_i}} \mathbf{R} \tag{2.16}$$

where $\mathbf{R}$ is a random vector with components drawn from a standard normal distribution, $T$ is the target temperature, and $k_\mathrm{B}$ is the Boltzmann constant. This procedure is known to sample from the correct thermodynamic ensemble and to preserve ergodicity. However, it also leads to decreased velocity time correlations, which in turn renders the Andersen thermostat unsuitable for the calculation of dynamical properties and transport coefficients.[141]

An alternative approach is adopted by the Bussi thermostat,[142] which scales the velocities of all particles by a common factor: $\dot{\mathbf{r}}_i' = \dot{\mathbf{r}}_i \left(\bar{K}/K\right)^{\frac{1}{2}}$. Here $K$ is the kinetic energy of the system prior to velocity scaling and $\bar{K} = N_f k_\mathrm{B} T/2$ is the target kinetic energy, which depends on the target temperature and the number of degrees of freedom, $N_f$. The kinetic energy is then propagated as an auxiliary dynamical variable according to

the stochastic differential equation

$$dK = (\bar{K} - K)\frac{dt}{\tau_T} + 2\sqrt{\frac{K\bar{K}}{N_f}}\frac{dW}{\sqrt{\tau_T}}, \tag{2.17}$$

where $\tau_T$ is a coupling constant and $dW$ denotes a Wiener process. In Equation 2.17, the first term causes an exponential relaxation of the kinetic energy towards its target value, while the second term ensures that the correct thermodynamic ensemble is sampled. Due to the stochastic nature of the evolution of $K$, this method is also known as stochastic velocity rescaling thermostat. In contrast to the Andersen thermostat, it has been shown not to interfere with the dynamical properties of a simulation system.[141]

Pressure can be controlled with a Parrinello-Rahman barostat[143,144] by dynamically evolving the size of a simulation cell. Since a protein-bilayer system may react differently to lateral stresses than to stresses applied in the direction perpendicular to the membrane, it is necessary to allow each box dimension to vary independently. To this end, the simulation cell is parameterised through a $3 \times 3$ matrix $\mathbf{B}$, in which each column corresponds to one of the edge vectors spanning the cell. This box matrix is then updated according to

$$\ddot{\mathbf{B}} = V\mathbf{W}^{-1}\mathbf{B}^{-\intercal}(\boldsymbol{\sigma} - \bar{\boldsymbol{\sigma}}), \tag{2.18}$$

where $V = \det(\mathbf{B})$ is the box volume, $\mathbf{W}^{-1}$ is a matrix of coupling parameters, and $\boldsymbol{\sigma}$ and $\bar{\boldsymbol{\sigma}}$ denote the instantaneous and target stress tensor respectively. Note that the scalar hydrostatic pressure is given by $P = \frac{1}{3}\operatorname{Tr}(\boldsymbol{\sigma})$.

Variation of box size introduces a volumetric contribution into the Hamiltonian of the system, which leads to a modified equation of motion of the form

$$\ddot{\mathbf{r}}_i = \frac{\mathbf{F}_i}{m_i} - \left(\mathbf{B}^{-1}(\mathbf{B}\dot{\mathbf{B}}^{\intercal} + \dot{\mathbf{B}}\mathbf{B}^{\intercal})\mathbf{B}^{-\intercal}\right)\dot{\mathbf{r}}_i, \tag{2.19}$$

which differs from Equation 2.1 by a fictitious friction term. The elements of the coupling matrix in Equation 2.18 can be expressed as

$$(\mathbf{W}^{-1})_{ij} = \frac{4\pi^2 \beta_{ij}}{3\tau_P^2 \max_{i,j}(B_{ij})} \tag{2.20}$$

in terms of the isothermal compressibilities, $\beta_{ij}$, and a coupling time constant, $\tau_P$. In practice, it is common to use the compressibility of water under ambient conditions of $\beta_{\mathrm{H_2O}} = 4.5 \times 10^{-5}\,\mathrm{bar}$.

An alternative method is the Monte Carlo barostat,[145,146] which employs a rejection sampling procedure to generate stochastic box volume changes. A trial volume change, $\Delta V$, is drawn from a uniform distribution over the interval $[-V_{\max}, V_{\max}]$ and the edge length of the simulation box is scaled by the relative volume change, i.e.

$$L'_\alpha = \frac{V + \Delta V}{V} L_\alpha, \tag{2.21}$$

where $\alpha \in \{x, y, z\}$ denotes the spatial dimension. By scaling box dimensions independently, the Monte Carlo barostat can achieve an anisotropic stress response appropriate for membrane protein simulations. Particle positions are similarly rescaled according to

$$r'_\alpha = \frac{V + \Delta V}{V}(r_\alpha - c_\alpha) + c_\alpha, \tag{2.22}$$

where $r_\alpha$ and $c_\alpha$ denote the $\alpha$-th element of the position vector $\mathbf{r}$ and box centre $\mathbf{c}$, where the particle index has been omitted for notational clarity.

The thermodynamic cost of the proposed volume change is then calculated as

$$\Delta\tilde{H} = \Delta U + P\Delta V - N k_\mathrm{B} T \ln\left(\frac{V + \Delta V}{V}\right), \tag{2.23}$$

where $\Delta U$ is the change in potential energy resulting from the change of volume, while $P$, $N$, and $T$ denote pressure, number of particles, and temperature respectively. The

quantity $\Delta\tilde{H}$ takes the form of a thermodynamic potential and bears similarity to the canonical enthalpy. Evaluating Equation 2.23 is computationally costly, as it necessitates a recalculation of the potential energy. Volume changes are therefore only attempted at every $n$-th time step, where in practice $n \approx 20$.

A proposed volume change is accepted with probability

$$P(\Delta\tilde{H}) = \begin{cases} 1 & \text{if } \Delta\tilde{H} > 0 \\ \exp\left(-\frac{\Delta\tilde{H}}{k_\mathrm{B}T}\right) & \text{if } \Delta\tilde{H} \leq 0 \end{cases} \tag{2.24}$$

and rejected otherwise. The parameter $V_\mathrm{max}$ is dynamically adjusted during the simulation to achieve an acceptance probability of approximately 50 %. This procedure ensures sampling from the isothermal-isobaric ensemble, provided that temperature is maintained at the appropriate value, which is commonly achieved through an Andersen thermostat.

### Periodic Boundaries and Long-Ranged Forces

In practice, the number of particles that can be simulated is limited by the available computational resources. To avoid a system boundary with a vacuum, periodic boundary conditions are commonly employed. In this approach, a finite simulation cell is surrounded by an infinite number of identical periodic images of itself and particles leaving a periodic cell re-enter it through the opposite boundary. As a consequence of this, angular momentum is not conserved in MD simulations.

A further complication arises in the treatment of long-ranged forces and in particular the electrostatic interaction. Under periodic boundary conditions, the total Coulomb potential of the simulation system is given by

$$U_\mathrm{ele} = \sum_{\mathbf{n}\in\mathbb{Z}^3}{}' \sum_{i,j} \frac{q_i q_j}{r_{ij,\mathbf{n}}}, \tag{2.25}$$

where $\mathbf{n} = (n_x, n_y, n_z)^\intercal$ runs over all copies of the simulation cell while the inner sum runs over all particles in a given periodic image and the prime indicates that for $\mathbf{n} = \mathbf{0}$ the case where $i = j$ needs to be omitted. The series in Equation 2.25 is only conditionally convergent and unsuitable for numerical evaluation.[147]

A more efficient algorithm known as Ewald summation[148] is obtained by decomposing the potential into a short-range component evaluated in direct space, $U_{\text{dir}}$, a long-range component evaluated in reciprocal space, $U_{\text{rec}}$, and a self-interaction term, $U_{\text{sel}}$. Using this formalism, the total electrostatic potential is given by

$$U_{\text{ele}} = U_{\text{dir}} + U_{\text{rec}} + U_{\text{sel}} \tag{2.26}$$

$$U_{\text{dir}} = \frac{1}{2} \sum_{\mathbf{n}\in\mathbb{Z}^3}{}' \sum_{i,j} q_i q_j \frac{\operatorname{erfc}\left(\kappa\, r_{ij,\mathbf{n}}\right)}{r_{ij,\mathbf{n}}} \tag{2.27}$$

$$U_{\text{rec}} = \frac{1}{2\pi V} \sum_{i,j} q_i q_j \sum_{\mathbf{k}\neq\mathbf{0}} \frac{\exp\left(-(\pi\kappa\mathbf{k})^2 + 2\pi i \mathbf{k}\cdot\mathbf{r}_{ij}\right)}{\mathbf{k}^2} \tag{2.28}$$

$$U_{\text{sel}} = -\frac{\kappa}{\sqrt{\pi}} \sum_i q_i^2, \tag{2.29}$$

where $V$ is the volume of the periodic cell, $\mathbf{k} = (k_x, k_y, k_z)^\intercal$ denotes the wave vector, and $\kappa$ is known as Ewald parameter.

Both the direct and reciprocal space contributions are exponentially convergent with respect to $\mathbf{n}$ and $\mathbf{k}$ respectively so that both series can be truncated without significant loss of accuracy.[149] For sufficiently large $\kappa$, only the leading term of the direct space sum ($\mathbf{n} = \mathbf{0}$) is non-negligible. Consequently, only the contributions of particles within a cutoff distance $r_{\text{cut}}$ need to be considered. For the CHARMM36m force field the appropriate cutoff value is $r_{\text{cut}} = 1.2\,\text{nm}$.[109] The self-interaction term is a finite sum whose evaluation is only of linear complexity in the number of particles, leaving the reciprocal space sum as the primary computational bottleneck.

For a system of $n$ particles, evaluation of the reciprocal space sum in Equation 2.28 has a time complexity of $\mathcal{O}(n^{3/2})$.[150] However, a complexity of $\mathcal{O}(n \log n)$ can be achieved

with the particle mesh Ewald (PME) method.[151] This is accomplished by smearing out the discrete particle charges onto a regular grid and solving the Poisson equation for the electrostatic potential. The Poisson solver can then utilise the Fast Fourier Transform (FFT) and electrostatic forces can be obtained by interpolating the derivative of the potential back onto the particle positions. In the smooth particle mesh Ewald (SPME) method,[152] discrete charges are mapped onto the grid through B-spline interpolation, which subsequently permits an analytical evaluation of the potential derivative.

Both GROMACS and OpenMM make use of the SPME method. In OpenMM, the Ewald parameter is specified implicitly through $\kappa = \sqrt{-\log(2\delta)}\, r_{\text{cut}}^{-1}$, where $\delta$ is a user-defined error tolerance. The elements of the maximum wave vector, $\mathbf{k}^{\max} = (k_x^{\max}, k_y^{\max}, k_z^{\max})^\intercal$, are then chosen as the smallest integers that fulfil

$$\delta > \frac{k_i^{\max}\sqrt{L_i\kappa}}{20} \exp\left(-\frac{\pi k_i^{\max}}{L_i\kappa}\right), \tag{2.30}$$

where $i \in \{x, y, z\}$ and $L_i$ is the length of the simulation cell along the $i$-th dimension. In GROMACS, $\mathbf{k}^{\max}$ is chosen by the user directly.

Van der Waals interactions decay much faster than the Coulomb force so that their effect on particle acceleration can simply be neglected if the pairwise distance exceeds a cutoff. In practice, the Lennard-Jones potential is smoothly decreased to zero between a switching distance, $r_{\text{swi}}$, and the cutoff distance, $r_{\text{cut}}$, through scaling with

$$s(x) = \begin{cases} 1 & \text{if } r_{ij} < r_{\text{swi}} \\ 1 - 6x^5 + 15x^4 - 10x^3 & \text{if } r_{\text{swi}} \leq r_{ij} \leq r_{\text{cut}} \\ 0 & \text{if } r_{ij} > r_{\text{cut}} \end{cases} \tag{2.31}$$

where $x = (r - r_{\text{swi}})/(r_{\text{cut}} - r_{\text{swi}})$.

However, for simulations at constant pressure, dispersion interactions have a significant influence on how the energy of a system varies with volume changes.[153] This is taken

into account by amending the internal energy with a long-range dispersion correction,

$$U_{\mathrm{cor}} = \frac{8\pi N}{V}\left(\frac{\sum_{i<j}\epsilon_{ij}\sigma_{ij}^{12}}{9r_{\mathrm{cut}}^{-9}} - \frac{\sum_{i<j}\epsilon_{ij}\sigma_{ij}^{6}}{3r_{\mathrm{cut}}^{-3}}\right), \tag{2.32}$$

where $N$ is the number of particles in the system and the summation goes over all pairwise interactions. A similar correction is applied to the pressure.[147]

## 2.2 Polarisation Effects in Classical Molecular Dynamics

Pairwise additive potentials underlie the predominant force fields currently employed in MD simulations of biomolecules and their accuracy has continuously improved over the course of their development.[154] However, the transferability of their parameters is in part limited by their inability to model the effect of induced polarisation. To address this problem, three approaches for incorporating polarisation effects into classical MD simulations have been proposed: fluctuating charge models permit atomic partial charges to vary over time, Drude oscillator models represent electronic degrees of freedom through additional particles, and polarisable atomic multipole models introduce inducible point dipoles as per-atom parameters.[155] While fluctuating charge models suffer from theoretical shortcomings[†], the other two approaches are actively pursued and have led to the development of new polarisable force fields. They are described below together with the additional algorithms that are required to evaluate dipole moments in an MD simulation.

### 2.2.1 The Classical Drude Oscillator

In the classical Drude oscillator approach, an atom is modelled as two individual mobile particles: a massive nucleus and a massless Drude oscillator, which represents the electronic degrees of freedom. The net charge of the atom, $q_{\mathrm{A}} = q_{\mathrm{N}} + q_{\mathrm{D}}$, is split between the charges of the nucleus, $q_{\mathrm{N}}$, and the Drude oscillator, $q_{\mathrm{D}}$, so that a displacement, $\mathbf{d}$, of

[†] A simple fluctuating charge model is unable to describe out-of-plane polarisation in e.g. aromatic rings.

the Drude oscillator with respect to the nucleus gives rise to an atomic dipole moment.

The interaction between a Drude oscillator and the corresponding nucleus is modelled as a harmonic spring by introducing a self-polarisation term of the form

$$U_{\text{self}} = \frac{1}{2} k_{\text{D}} \mathbf{d}^2 \tag{2.33}$$

into the potential energy. Here $k_{\text{D}}$ is a force constant which combined with $q_{\text{D}}$ determines the polarisability

$$\alpha = \frac{q_{\text{D}}^2}{k_{\text{D}}} \tag{2.34}$$

of the atom.

Outside of the self-polarisation term, the Drude oscillators contribute to the potential energy of the system only through Coulomb interactions with other charge carrying particles. They are not affected by dispersion-repulsion type interactions and do not participate in any bonded interactions.

Various sets of Drude polarisation parameters for water[156,157] and ionic solutions[158,159] have been proposed in the literature. However, the most mature force field based on the classical Drude oscillator is the CHARMM polarisable force field,[160] which provides consistent parameters for all constituents of a typical biomolecular simulation, notably for water, ions, proteins, and lipids. This force field has been implemented in a variety of common simulation packages, including CHARMM,[161] GROMACS,[100] NAMD,[162] LAMMPS,[163] and OpenMM.[164] Its details are discussed in the following section.

#### The CHARMM Polarisable Force Field

Development of the CHARMM Drude force field began with the creation of the polarisable SWM4-DP model of water,[165] which comprises four regular interaction sites similar to the TIP4P water model plus an additional Drude oscillator attached to the oxygen

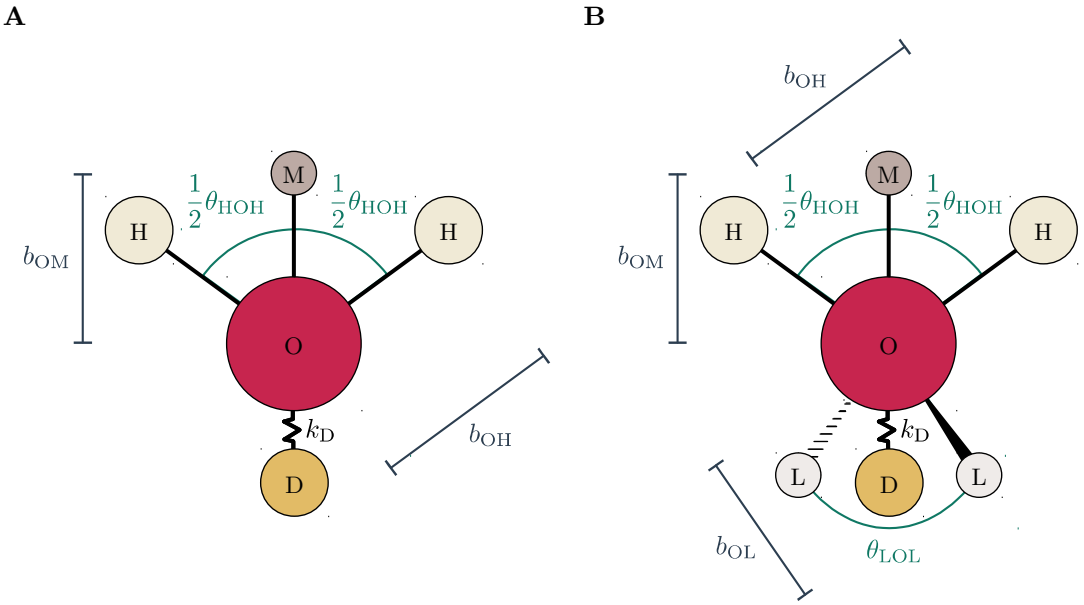



**Figure 2.5:** Geometry of CHARMM Drude water models. (A) The SWM4-DP and SWM4-NDP water models share the basic geometry of the TIP4P additive models with a single van der Waals interaction site ($\epsilon_O$, $R_O$) on the oxygen atom (red) and charges $q_H$ and $q_M$ located on the hydrogens (white) and a virtual "M"-site (light brown) respectively. A Drude oscillator (yellow) is attached to the oxygen atom via a harmonic spring with force constant $k_D$ and both of these particles carry equal but opposite charges ($q_O$, $q_D = -q_O$), so that the net charge of the Drude-oxygen pair is zero. For visual clarity, the D–O distance is exaggerated in this schematic. In practice, the Drude oscillator never leaves the van der Waals sphere of the oxygen atom and without an external electric field their resting distance is zero. (B) The SWM6 model adds two virtual "L"-sites (light grey) to the geometry of the SWM4 models. These sites represent lone pair electrons and together with the two hydrogens form a tetrahedral arrangement around the central oxygen. Sizes and distances are not drawn to scale, see Table 2.2 for the geometric parameters of the CHARMM Drude water models.

atom (Figure 2.5). In this model, the Drude particle carries a positive charge and it has since been superseded by the SWM4-NDP water model,[166] a reparameterisation with a negatively charged Drude oscillator. More recently, an additional six site model called SWM6[167] was proposed, where two additional virtual sites are used to represent the lone pair electrons on the oxygen. The parameters of these water models are summarised in Table 2.2.

A consistent parameterisation of polypeptides has been achieved by the CHARMM Drude 2013 protein force field.[168] Its functional form is similar to that of the additive

| Parameter | Units | SWM4-DP[165] | SWM4-NDP[166] | SWM6[167] |
|---|---|---|---|---|
| $b_{\mathrm{OH}}$ | nm | 0.9572 | 0.0957 | 0.0957 |
| $b_{\mathrm{OM}}$ | nm | 0.2381 | 0.0240 | 0.0247 |
| $b_{\mathrm{OL}}$ | nm | — | — | 0.0315 |
| $\theta_{\mathrm{HOH}}$ | ° | 104.52 | 104.52 | 104.52 |
| $\theta_{\mathrm{LOL}}$ | ° | — | — | 101.10 |
| $k_{\mathrm{D}}$ | $\mathrm{MJ/mol/nm^2}$ | 418.40 | 418.40 | 418.40 |
| $q_{\mathrm{O}}$ | $e$ | −1.7719 | 1.7164 | 1.9159 |
| $q_{\mathrm{D}}$ | $e$ | 1.7719 | −1.7164 | −1.6279 |
| $q_{\mathrm{H}}$ | $e$ | 0.5537 | 0.5573 | 0.5307 |
| $q_{\mathrm{M}}$ | $e$ | −1.1074 | −1.1147 | −1.1334 |
| $q_{\mathrm{L}}$ | $e$ | — | — | −0.1080 |
| $\epsilon_{\mathrm{O}}$ | $\mathrm{kJ\,mol^{-1}}$ | 0.8606 | 0.8826 | 0.6778 |
| $\sigma_{\mathrm{O}}$ | nm | 0.1417 | 0.1592 | 0.1599 |

**Table 2.2:** Intramolecular, electrostatic, and van der Waals parameters of the CHARMM Drude water models. See Figure 2.5 for the definition of these parameters.

CHARMM force field, but its parameters have been reoptimised to match data from quantum mechanical calculations as well as from NMR measurements of oligopeptide dynamics. A negatively charged Drude oscillator is attached to each heavy atom, while hydrogens are treated as non-polarisable.

In addition to Drude oscillators, the force field prescribes lone pair virtual sites on some heavy atoms to better account for the anisotropic distribution of electronic charges. In conjunction with the introduction of these virtual sites, the CHARMM Drude 2013 force field also allows for anisotropic polarisability. To achieve this, the self-polarisation energy in Equation 2.33 is replaced by

$$U_{\mathrm{self}} = \frac{1}{2}\,\mathbf{d}\cdot\mathbf{K}\cdot\mathbf{d}, \tag{2.35}$$

where $\mathbf{K}$ is a second rank tensor representing an anisotropic spring constant. In the

optimisation of the force field parameters, $\frac{1}{3}\operatorname{Tr}\mathbf{K} = 418\,\mathrm{MJ/mol/nm^2}$ is enforced, so that the atomic polarisability is controlled solely through the Drude oscillator charge $q_\mathrm{D}$ and $\mathbf{K}$ modulates the anisotropy, but not the magnitude, of the polarisability tensor, $\boldsymbol{\alpha}$. Anisotropy is particularly important for electronegative atoms and has been shown to improve the description of hydrogen bonding.

A further important difference between the additive and polarisable CHARMM force fields lies in the treatment of electrostatic interactions within a molecule. In the additive force field, Coulomb forces are not calculated between nearest (1–2 pairs) and next-nearest (1–3 pairs) neighbours along a chain of bonds, as their mutual interaction is considered to be fully described by the bond stretching and angle bending terms of the force field. With the introduction of explicit atomic polarisation this mean field treatment is no longer valid, as the dipole moments of bonded atoms will influence one another.

To account for these dipole-dipole interactions, the Coulomb force between the Drude oscillators of 1–2 and 1–3 pairs are evaluated explicitly, but are scaled down by an electrostatic screening factor. Coulomb forces between nuclei and Drude oscillators (charge-dipole interactions) as well as between two nuclei (charge-charge interactions) are not evaluated for 1–2 and 1–3 pairs, while all three interaction types are evaluated for 1–4 and higher order pairs. The screening factor applied to the dipole-dipole interactions,

$$S_{ij} = 1 - \left(1 + \frac{a\tilde{r}_{ij}}{2}\right)\exp\left(-a\tilde{r}_{ij}\right), \tag{2.36}$$

was derived by Thole.[169] It is based on smearing out the discrete point charges into a continuous density,

$$\rho = \frac{a^3}{8\pi}\exp\left(-a\tilde{r}_{ij}\right), \tag{2.37}$$

where $\tilde{r}_{ij} = r_{ij}/(\alpha_i\alpha_j)^{1/6}$ is the interatomic distance, $\alpha_i = \operatorname{Tr}\boldsymbol{\alpha}_i$ represents the polaris-

ability of the $i$-th atom, and $a = a_i + a_j$ is the sum of the atom specific Thole screening parameters. Thole screening is also applied to some non-bonded interactions, notably those involving divalent ions and in this case $S_{ij}$ is also used to scale charge-dipole and charge-charge interactions.

A lipid parameter set was first developed for dipalmitoylphosphatidylcholine (DPPC)[170] and has since been extended to include a variety of lipids with phosphatidylcholine (PC) and phosphatidylethanolamine (PE) head groups in combination with both saturated and monounsaturated hydrocarbon tails.[171] As in the protein force field, Drude oscillators are only associated with heavy atoms and hydrogens are non-polarisable. It is worth noting that the CHARMM polarisable force field is also available for DNA[172,173] and RNA[174] as well as for carbohydrates[175,176] and molecular ions,[177] although none of these parameters were used in this thesis.

### Molecular Dynamics Algorithms for Drude Oscillators

Using the force field described above in an MD simulation requires an equation of motion for propagating the positions of the nuclei and a procedure for updating the relative displacements of the Drude oscillators. In the basic formulation of the Drude model, the oscillators are massless and only the positions of the nuclei, $\mathbf{r}_i$, are treated as dynamical variables. A self-consistent field (SCF) approach is employed to determine the Drude displacements that minimise the potential energy for fixed nucleus positions by solving

$$\nabla_{\mathbf{d}_i} U = \nabla_{\mathbf{d}_i} \left( U_{\text{self}} + U_{\text{Coulomb}} \right) = \mathbf{0} \tag{2.38}$$

for $\mathbf{d}_i$, where $i$ runs over all polarisable atoms. The potential solving Equation 2.38, $U^{\text{SCF}}$, is then used in the equation of motion for the nuclei,

$$m_i \ddot{\mathbf{r}}_i = -\nabla_{\mathbf{r}_i} U^{\text{SCF}}, \tag{2.39}$$

which can be integrated through the same methods as the regular Newton equation.

The SCF approach directly reflects the physical assumption underlying the Born-Oppenheimer approximation that the electronic degrees of freedom relax on a much shorter time scale than any nuclear motion and that the electrons in an MD simulation are therefore always in the lowest energy state. In practice however, Equation 2.38 must be solved iteratively, necessitating a repeated evaluation of the Coulomb interactions. As this incurs a prohibitive computational cost, a more efficient dual-thermostat extended Lagrangian[161] scheme is commonly used.

In this approach, each Drude oscillator is associated with a small mass, $m_\mathrm{D}$, of typically 0.4 u, which is subtracted from the mass of the nucleus to preserve the total mass of the atom, $m_i$. The reduced mass $m_i' = m_\mathrm{D}(1-m_\mathrm{D}/m_i)$ can then be interpreted as a moment of inertia so that the Drude displacements can be integrated as additional dynamical variables. It has been shown that to reproduce accurately the dynamics observed in the SCF regime, the Drude oscillators need to be maintained at a fictitious temperature, $T_\star \ll T$, which is achieved by coupling nucleus positions and Drude displacements to two separate thermostats.

The exact choice of thermostat varies between MD engines. While CHARMM and GROMACS employ a dual Nosé-Hoover thermostat,[178,179] NAMD and OpenMM implement dual Langevin dynamics.[147] In the latter formalism, the coupled equations of motion take the form

$$\begin{aligned} m_i\ddot{\mathbf{r}}_i &= \nabla_{\mathbf{r}_i}U - \gamma\dot{\mathbf{r}}_i + \sqrt{2\gamma k_\mathrm{B}T}\mathbf{R}(t) \\ m_i'\ddot{\mathbf{d}}_i &= \nabla_{\mathbf{d}_i}U - \gamma_\star\dot{\mathbf{d}}_i + \sqrt{2\gamma_\star k_\mathrm{B}T_\star}\mathbf{R}_\star(t) \end{aligned} \tag{2.40}$$

where $\gamma$ and $\gamma_\star$ are friction coefficients. $\mathbf{R}(t)$ and $\mathbf{R}_\star(t)$ are zero mean delta-correlated Gaussian processes, which represent the coupling of the dynamic variables to a heat bath of temperature $T$ and $T_\star$ respectively. Equation 2.40 can be solved through an extension of the Verlet method.[180]

When the extended Lagrangian scheme is used to update Drude particle positions, the frequency of the harmonic oscillation associated with the self-polarisation energy in Equation 2.35 replaces the vibration of covalently bonded hydrogens as the step size limiting time scale. In addition, large displacements of the Drude particle may incur instabilities known as polarisation catastrophe. To remedy this, a hard-wall constraint[170] is applied after the position update step of the Verlet integrator. If a Drude displacement exceeds a predefined threshold ($|\mathbf{d}_i| > d_{\mathrm{hw}}$), the Drude particle will be reflected back towards the nucleus. In order to preserve linear momentum, the velocity of the atom will be adjusted accordingly.

### 2.2.2 Polarisable Atomic Multipoles

An alternative approach to modelling electronic polarisation in a classical MD setting is the inclusion of higher order multipole moments in the calculation of electrostatic interactions. This is motivated by the fact that atomic orbitals can mathematically be described by a multipole expansion. In most commonly used force fields, including the CHARMM36m and CHARMM Drude force fields, this multipole expansion is truncated at the monopole order and the Coulomb potential is used to describe charge-charge interactions. While this greatly simplifies the calculation of electrostatic forces, the inclusion of higher order multipole moments is expected to give a more accurate representation of the true distribution of electronic charge around an atom.

The atomic multipole approach to induced polarisation does not rely on additional particles and instead each atom is directly associated with a multipole moment vector, $\mathbf{M}_i$, and the Coulomb potential for the electrostatic interaction between atoms $i$ and $j$ is replaced by

$$U_{\mathrm{ele}} = \mathbf{M}_i^{\intercal} \mathcal{T}_{ij} \mathbf{M}_j, \tag{2.41}$$

where $\mathcal{T}_{ij}$ is a multipole-multipole interaction tensor.[181,182] The multipole moment vector is defined with respect to an atom-centred reference frame, with basis vectors defined by the relative position of neighbouring atoms. It can be decomposed into a permanent contribution, $\mathbf{M}_i^{\text{per}}$, which is independent of the molecular environment, and an induced component, $\mathbf{M}_i^{\text{ind}}$, which adapts to the local electric field caused by other atoms.[183]

The induced multipole moment is proportional to the electric field arising from both the permanent and induced multipole moments and is given by

$$M_{i,\alpha}^{\text{ind}} = \alpha_i \left( \sum_j \mathbf{T}_{ij}^{\alpha} \mathbf{M}_j^{\text{per}} + \sum_{k \neq i} \mathbf{T}_{ik}^{\alpha} \mathbf{M}_k^{\text{ind}} \right), \tag{2.42}$$

where the proportionality constant $\alpha_i$ represents an atomic polarisability similar to that defined in Equation 2.34 for the Drude force field. The components of the interaction vector

$$\mathbf{T}_{ij}^{\alpha} = \begin{pmatrix} T_{ij}^{\alpha} & T_{ij}^{\alpha 1} & T_{ij}^{\alpha 2} & T_{ij}^{\alpha 3} & T_{ij}^{\alpha 11} & T_{ij}^{\alpha 12} & \dots \end{pmatrix} \tag{2.43}$$

are recursively defined according to

$$T_{ij} = \frac{1}{r_{ij}} \qquad T_{ij}^{\alpha} = \frac{\partial}{\partial r_i^{\alpha}} T_{ij} \qquad T_{ij}^{\alpha\beta} = \frac{\partial}{\partial r_i^{\alpha}} T_{ij}^{\beta} \qquad T_{ij}^{\alpha\beta\gamma} = \frac{\partial}{\partial r_i^{\alpha}} T_{ij}^{\beta\gamma} \tag{2.44}$$

where the derivatives are taken with respect to the $\alpha$-th component of the position vector $\mathbf{r}_i = (r_i^x \, r_i^y \, r_i^z)^{\intercal}$ of the $i$-th atom.[184]

A consistent parameterisation of polarisable multipole interactions for the simulation of biomolecules is the Atomic Multipole Optimised Energetics for Biomolecular Applications (AMOEBA) force field,[99] which comprises parameters for water,[185,186] atomic ions,[187,188] small organic molecules,[189] proteins,[190] and nucleic acids.[191] Parameters for lipids are not officially included in the AMOEBA force field, but have been proposed in the literature.[192,193] Next to its reference implementation in Tinker,[194] the AMOEBA

force field is available in Amber,[195] OpenMM,[196] and Tinker-HP.[197] The following section will give a brief overview of this force field.

### The AMOEBA Force Field

The AMOEBA force field did not evolve out of an additive force field and its functional form has been developed from the ground up within the polarisable multipole framework outlined above. Its internal energy $U = U_{\text{bnd}} + U_{\text{vdW}} + U_{\text{ele}}$ comprises contributions from covalent bonds, $U_{\text{bnd}}$, dispersion-repulsion interactions, $U_{\text{vdW}}$, and the electrostatic multipole potential from Equation 2.41.

The contribution of covalent bonds is given by

$$U_{\text{bnd}} = U_b + U_\theta + U_{b\theta} + U_\chi + U_\phi \tag{2.45}$$

$$U_b = k_b \left(b - b_0\right)^2 \left[1 - a_b(b - b_0) + b_b(b - b_0)^2\right] \tag{2.46}$$

$$U_\theta = k_\theta(\theta - \theta_0)^2 \left[1 - a_\theta(\theta - \theta_0) + b_\theta(\theta - \theta_0)^2 - c_\theta(\theta - \theta_0)^3 + d_\theta(\theta - \theta_0)^4\right] \tag{2.47}$$

$$U_{\text{b}\theta} = k_{\text{b}\theta}(\theta - \theta_0)\left[(b - b_0) + (b' - b'_0)\right] \tag{2.48}$$

$$U_\chi = k_\chi \chi^2 \tag{2.49}$$

$$U_\phi = \sum_n k_{n\phi}\left[1 + \cos(n\phi \pm \delta)\right] \tag{2.50}$$

where $U_b$ and $U_\theta$ are the energy of stretching bond length, $b$, and bending bond angle, $\theta$, relative to the equilibrium length, $b_0$, and equilibrium angle, $\theta_0$, while $U_{b\theta}$ is a joint stretch-bending term. All three terms are described as anharmonic springs with force constants $k_b$, $k_\theta$, and $k_{b\theta}$ respectively and the values of the anharmonicity parameters are $a_b = 2.55$ and $b_b = 1.4874$ for bond stretching and $a_\theta = 0.014$, $b_\theta = 5.6 \times 10^{-5}$, $c_\theta = 7.0 \times 10^{-7}$, and $d_\theta = 2.2 \times 10^{-8}$ for angle bending. Out of plane bending is modelled through a Wilson-Decius-Cross[198] term, $U_\chi$, where $\chi$ is the angle of a bond relative to a plane defined by three sequentially bonded atoms, and $k_\chi$ is the corresponding force constant. Finally, backbone torsion is described through $U_\phi$, where $\phi$ is the dihedral

angle and $\delta$ is a phase shift.

Instead of employing the Lennard-Jones potential, van der Waals interactions are modelled through a buffered 14–7 potential[199] of the form

$$U_{\mathrm{vdW}} = \epsilon_{ij} \left( \frac{1+\delta}{\rho_{ij}+\delta} \right)^7 \left( \frac{1+\gamma}{\rho_{ij}^7+\gamma} - 2 \right) \tag{2.51}$$

with parameters $\delta = 0.07$ and $\gamma = 0.12$. Here $\epsilon_{ij}$ is the depth of the potential well and $\rho_{ij} = r_{ij}/R_{ij}$ is the distance $r_{ij}$ between the $i$-th and $j$-th atom in units of the distance of minimal energy, $R_{ij}$. If atoms $i$ and $j$ are of different types, the combining rules

$$R_{ij} = \frac{(R_{ii})^3 + (R_{jj})^3}{(R_{ii})^2 + (R_{ij})^2} \qquad \text{and} \qquad \epsilon_{ij} = \frac{4\epsilon_{ii}\epsilon_{jj}}{\left(\sqrt{\epsilon_{ii}} + \sqrt{\epsilon_{jj}}\right)^2} \tag{2.52}$$

apply, where $R_{ii}$ and $\epsilon_{ii}$ are atom specific force field parameters.

The electrostatic potential is given in Equation 2.41, where permanent multipole moments up to quadrupole order are included. Therefore, the AMOEBA multipole vector can be expressed as

$$\mathbf{M}_i = \begin{pmatrix} q_i & \mu_i^x & \mu_i^y & \mu_i^z & Q_i^{xx} & Q_i^{xy} & Q_i^{xz} & \dots & Q_i^{zz} \end{pmatrix}^\intercal \tag{2.53}$$

where $q_i$ is the partial charge, $\mu_i^\alpha$ are the components of the point dipole vector, and $Q_i^{\alpha\beta}$ are the components of the point quadrupole tensor at the position of the $i$-th atom. However, only the dipole moment is treated as inducible, while charge and quadrupole moment are invariable. The current AMOEBA force field is restricted to isotropic polarisabilities, $\alpha_i$, while the CHARMM Drude force field permits anisotropy through the polarisability tensor $\boldsymbol{\alpha}_i$. A further difference is that AMOEBA treats all hydrogens as polarisable.

In the calculation of induced dipoles, the AMOEBA model prescribes a special treatment for intramolecular polarisation. While the second sum on the right hand side of

Equation 2.42 runs over all other atoms ($k \neq i$), the first sum excludes all atoms within the same polarisation group as atom $i$. For small molecules, the polarisation group comprises the entire molecule, while for larger molecules polarisation groups roughly correspond to functional groups and are chosen to ensure limited intra-group conformational flexibility.

In order to prevent the divergence of induced dipole moments at short interatomic distances, the AMOEBA force field employs a version of Thole screening adapted for multipoles.[200] In this approach, one of the point dipoles in a dipole-dipole interaction is represented by a spatially distributed charge density

$$\rho = \frac{3a}{4\pi} \exp\left(-a\tilde{r}_{ij}^3\right), \tag{2.54}$$

where $\tilde{r}_{ij} = r_{ij}/(\alpha_i\alpha_j)^{1/6}$ and $a$ are the reduced interatomic distance and the dimensionless Thole screening parameter previously defined. This is realised through a scaling factor of

$$S_{ij} = 1 - \exp\left(-a\tilde{r}_{ij}^3\right) \tag{2.55}$$

which is incorporated into the elements of the multipole interaction tensor $\mathcal{T}_{ij}$. However, the factor $S_{ij}$ is only applied to interactions involving induced dipoles, while interactions between permanent multipole moments remain unscaled. This treatment of Thole screening is conceptually similar to that employed in the Drude oscillator framework (Equation 2.36), but with a different underlying charge density. As a further distinction, the CHARMM Drude force field allows atom specific Thole factors $a_i$, while the AMOEBA force field adopts a universal value of $a = 0.39$.

Several explicit water models for simulations with the AMOEBA force field have been developed (Figure 2.6). The AMOEBA03 model[185] was the first parameterisation of a molecule in the AMOEBA framework and subsequently served as the host water

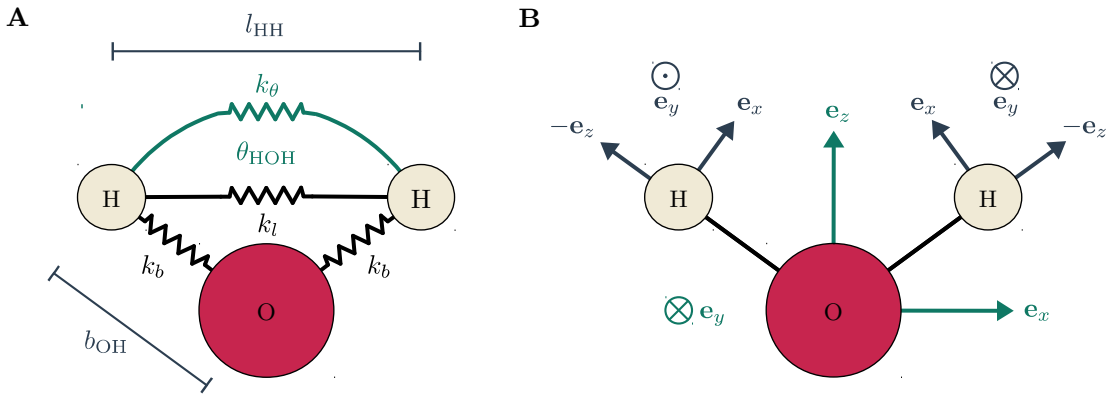



**Figure 2.6:** Geometry of the AMOEBA03 and AMOEBA14 water models. (A) The AMOEBA water models comprise three interaction sites located at the positions of the hydrogen (white) and oxygen (red) atoms of the water molecule. All atoms interact via both dispersion-repulsion and electrostatic multipole forces and are associated with van der Waals parameters ($\epsilon_{\mathrm{O}}$, $R_{\mathrm{O}}$, $\epsilon_{\mathrm{H}}$, $R_{\mathrm{H}}$) as well as with a multipole vector ($\mathbf{M}_{\mathrm{O}}$, $\mathbf{M}_{\mathrm{H}}$). The model geometry is fully flexible with the O–H bond lengths and the H–O–H bond angle subjected to anharmonic potentials with force constants $k_b$ and $k_\theta$ respectively. An additional Urey-Bradley potential with force constant $k_l$ is applied to restrain the H–H distance. (B) Dipole and quadrupole moments are defined in a local reference frame centred on the respective atom. For oxygen, the first axis ($\mathbf{e}_z$) points along the H–O–H bisector, while for hydrogen, it is aligned with the O–H bond. An orthogonal direction in the plane of the water molecule is taken as the second axis ($\mathbf{e}_x$) and the third axis ($\mathbf{e}_y$) is chosen to complete a right-handed orthonormal basis. Vectors oriented away from the reader are denoted with a $\otimes$, while vectors pointing at the reader are denoted by a $\odot$. Sizes and distances are not drawn to scale, see Table 2.3 for the geometric parameters of the AMOEBA water models.

model for the parameterisation of the AMOEBA protein force field. Significant improvements in reproducing the thermodynamic behaviour of real water were achieved by the AMOEBA14 water model,[186] which was developed using the force balance approach.[201] The parameters of these models are listed in Table 2.3.

Both models comprise three multipole and repulsion-dispersion interaction sites, where the centre of the van der Waals interaction of the hydrogens is shifted towards the oxygen by 9% and 20% of the bond length for AMOEBA03 and AMOEBA14 respectively. In contrast to the additive and Drude water models considered previously, AMOEBA water is fully flexible and the potential energy of the water model is amended by a harmonic Urey-Bradley term, $U_l = k_l(l - l_0)^2$, where $l$ is the distance between the water hydrogens

| Parameter | Units | AMOEBA03[185] | AMOEBA14[186] |
|---|---|---|---|
| $b_{\mathrm{OH}}$ | nm | 0.0957 | 0.0957 |
| $l_{\mathrm{HH}}$ | nm | 0.1533 | 0.1547 |
| $\theta_{\mathrm{HOH}}$ | ° | 108.50 | 107.91 |
| $k_{\mathrm{b}}$ | MJ/mol/nm$^2$ | 232.99 | 232.99 |
| $k_{\mathrm{l}}$ | MJ/mol/nm$^2$ | −3.18 | −3.61 |
| $k_\theta$ | kJ/mol/rad$^2$ | 203.80 | 204.93 |
| $q_{\mathrm{O}}$ | $e$ | −0.5197 | −0.4262 |
| $q_{\mathrm{H}}$ | $e$ | 0.2598 | 0.2131 |
| $\mu_z^{\mathrm{O}}$ | $e$ pm | 7.5561 | 3.3079 |
| $\mu_x^{\mathrm{H}}$ | $e$ pm | −2.0421 | −5.3537 |
| $\mu_z^{\mathrm{H}}$ | $e$ pm | −3.0788 | −14.3783 |
| $Q_{xx}^{\mathrm{O}}$ | $e$ pm$^2$ | 1062.09 | 492.18 |
| $Q_{yy}^{\mathrm{O}}$ | $e$ pm$^2$ | −1170.77 | −648.55 |
| $Q_{zz}^{\mathrm{O}}$ | $e$ pm$^2$ | 108.68 | 156.37 |
| $Q_{xx}^{\mathrm{H}}$ | $e$ pm$^2$ | −102.85 | 343.96 |
| $Q_{yy}^{\mathrm{H}}$ | $e$ pm$^2$ | −300.72 | 250.63 |
| $Q_{xy}^{\mathrm{H}}$ | $e$ pm$^2$ | −5.69 | −195.71 |
| $Q_{zz}^{\mathrm{H}}$ | $e$ pm$^2$ | 403.58 | −594.59 |
| $\alpha_{\mathrm{O}}$ | nm | 0.0837 | 0.0920 |
| $\alpha_{\mathrm{H}}$ | nm | 0.0496 | 0.0539 |
| $\epsilon_{\mathrm{O}}$ | kJ mol$^{-1}$ | 0.4600 | 0.6326 |
| $\epsilon_{\mathrm{H}}$ | kJ mol$^{-1}$ | 0.0565 | 0.0439 |
| $R_{\mathrm{O}}$ | nm | 0.1703 | 0.1790 |
| $R_{\mathrm{H}}$ | nm | 0.1328 | 0.1059 |

**Table 2.3:** Intramolecular, electrostatic, and van der Waals parameters of the AMOEBA water models. Components of the dipole and quadrupole moment not given in the table above are zero due to molecular symmetry. See Figure 2.6 for the definition of these parameters.

and $l_0$ is their equilibrium distance.

A further refinement of the water model was suggested recently[202] as a proof of principle for the inclusion of anisotropic polarisability into the AMOEBA force field. Simplified water models including the unified uAMOEBA[203] model, which describes each water molecule as a single interaction site, and the inexpensive iAMOEBA[204] model, based on neglecting mutual induction effects, have also been proposed.

### Molecular Dynamics Algorithms for Polarisable Atomic Multipoles

Applying the AMOEBA force field in an MD simulation necessitates the specification of an additional algorithm for determining the induced dipole moments for a given conformation of atomic positions in the simulation system. As in the case of the Drude oscillators, the electronic degrees of freedom are assumed to relax instantaneously and the self-consistency condition

$$\nabla_{\mathbf{M}_i^{\mathrm{ind}}} U = \mathbf{0} \tag{2.56}$$

is used to determine $\mathbf{M}_i$.

This is accomplished by reformulating Equation 2.42 into an implicit expression,

$$\mu_{i,\alpha}^{\mathrm{ind}}(n+1) = \mu_{i,\alpha}^{\mathrm{ind}}(n) - \alpha_i \sum_j T_{ij}^{\alpha,\beta} \mu_{j,\beta}^{\mathrm{ind}}(n), \tag{2.57}$$

for the induced atomic dipole moments, $\boldsymbol{\mu}_i^{\mathrm{ind}}$ (charges and quadrupole moments are non-inducible in the AMOEBA framework). Equation 2.57 can be solved iteratively by successive over-relaxation[185] or the direct inversion in iterative subspace algorithm,[205] but the number of iterations required for convergence to within an acceptable numerical tolerance can be prohibitively large.

One option is to neglect the effect of induced dipoles on one another and to only consider the field caused by permanent multipoles when calculating induced dipoles

(i.e. neglecting the second term on the right hand side of Equation 2.42). However, this "direct" polarisation approach reduces the accuracy of the force field and has been observed to cause dynamical instabilities. Instead, the extrapolated perturbation theory method,[206] which finds $\boldsymbol{\mu}_i^{\mathrm{ind}}(\infty)$ by extrapolating from the first few iterations of Equation 2.57, is used here. This method is significantly more efficient than a full self-consistency iteration, but still accounts for the mutual polarisation of induced dipoles.

A further step towards accelerating AMOEBA simulations is the application of the reversible reference system propagation algorithm[207] (r-RESPA) for solving the equation of motion. The r-RESPA method is a multiple time stepping integrator, in which the fast varying short-ranged interactions are evaluated more frequently than the slowly varying but computationally expensive long-ranged interactions. Multiple time stepping algorithms can also be combined with additive force fields, but they are especially beneficial for the AMOEBA force field, as they reduce the number of dipole self-consistency iterations per physical time interval.

## 2.3 Umbrella Sampling

An important quantity in the analysis of MD simulations is the potential of mean force (PMF), $\mathcal{W}(\boldsymbol{\xi})$, a configuration-dependent thermodynamic potential that describes the free energy cost of moving the simulation system along a set of $D$ collective variables $\boldsymbol{\xi} = \boldsymbol{\xi}(\{\mathbf{r}_i\}) = (\xi_1(\{\mathbf{r}_i\}), \xi_2(\{\mathbf{r}_i\}), \ldots, \xi_D(\{\mathbf{r}_i\}))^{\mathsf{T}}$. Of particular interest here is the 1D PMF of an ion passing through the pore of an ion channel, from which an estimate of channel conductance can be derived.[35] In this case, the collective variable is frequently taken to be the position[‡] of a single ion along the transmembrane axis ($\xi = z$), implicitly assuming that ion-ion correlations are negligible.

The PMF is related to the probability, $P(\xi)$, of finding the system at a given value of

[‡]Here it is assumed that the origin of the coordinate system is located at the centre of mass of the channel protein, as the collective variable should be the relative distance between ion and channel rather than their absolute position in the simulation cell.

the collective variable through the Boltzmann equation,

$$\mathcal{W}(\xi) = -k_{\mathrm{B}}T \ln P(\xi) + \mathcal{W}_0, \tag{2.58}$$

where $\mathcal{W}_0$ is an offset that does not influence the dynamics of the system and is usually chosen with respect to some reference state. Therefore, $P(\xi)$ can in principle be estimated directly from an MD trajectory. However, if $\mathcal{W}(\xi)$ contains high barriers, an MD simulation might not fully sample the entire range of $\xi$ within a computationally feasible simulation time.

An accurate estimate of the PMF can then be obtained via umbrella sampling.[208] In this approach, $N_{\mathrm{sim}}$ simulations are carried out, during each of which the force field potential is amended by a harmonic biasing potential,

$$w_i(\xi) = \frac{1}{2} k_\xi \left(\xi - \xi_i\right)^2, \tag{2.59}$$

with force constant $k_\xi$, which restrains the collective variable to values near $\xi_i$. This guarantees that the entire domain of $\xi$ is sampled. However, the probability distributions, $P_i(\xi)$, that can be estimated from these biased simulations, need to be combined into an overall distribution $P(\xi)$ in a way that corrects for the bias due to $w_i(\xi)$.

This can be accomplished by employing the weighted histogram analysis method (WHAM).[91,209] It is based on iteratively solving a coupled system of nonlinear equations,

$$P(\xi) = \frac{\sum_{i=1}^{N_{\mathrm{sim}}} n_i P_i(\xi)}{\sum_{j=1}^{N_{\mathrm{sim}}} n_j \exp\left(-\beta(w_j(\xi) - F_j)\right)} \tag{2.60}$$

and

$$F_j = -\beta \ln \left( \int d\xi \, \frac{P(\xi)}{\exp\left(\beta w_j(\xi)\right)} \right), \tag{2.61}$$

where $n_i$ is the total number of samples gathered from the $i$-th simulation, $\beta^{-1} = k_{\mathrm{B}}T$ denotes the thermodynamic temperature factor, and $F_j$ can be interpreted as the free energy required to introduce the $j$-th biasing potential into the system. In this thesis, solutions to Equations 2.60 and 2.61 were found using the `g_wham` tool in GROMACS[210] as well as the Grossfield lab standalone `wham` implementation.[211]

## 2.4 Continuum Electrostatics

While MD simulations provide atomistically detailed insights into biomolecular processes, they are computationally expensive. A simpler framework for the analysis of biomolecular solvation and electrostatics is provided by the Poisson-Boltzmann equation derived below.[212,213] Both protein and water are treated as dielectric continua and ionic mobility is incorporated into the theory through thermodynamic considerations. These approximations considerably improve computational efficiency, but Poisson-Boltzmann theory can not account for entropic effects.

Fundamentally, the electrostatic potential, $\phi(\mathbf{x})$, is related to the charge density, $\rho(\mathbf{x})$, through the Poisson equation,

$$-\nabla \cdot \big(\epsilon_0 \epsilon(\mathbf{x}) \nabla \phi(\mathbf{x})\big) = \rho(\mathbf{x}), \tag{2.62}$$

where $\epsilon(\mathbf{x})$ is the relative electric permittivity and $\epsilon_0$ denotes the permittivity of vacuum. The overall charge density comprises contributions from the solute, $\rho_{\mathrm{fix}}(\mathbf{x})$, and from the mobile ionic charges, $\rho_{\mathrm{mob}}(\mathbf{x})$, while effects of atomic and orientational polarisation are subsumed into the electric permittivity field. Typically, the permittivity of the solvent is taken to be that of bulk water, $\epsilon_{\mathrm{sol}} = 78$, while the solute permittivity ranges from $\epsilon_{\mathrm{pro}} = 2$, when only atomic polarisability is considered, up to $\epsilon_{\mathrm{pro}} = 20$, when sidechain flexibility is taken into account.

As in an MD force field, the solute is modelled as a spatial distribution of partial

charges, $q_i$, and the corresponding charge density is expressed as

$$\rho_{\text{fix}}(\mathbf{x}) = \sum_i q_i \delta(\mathbf{x} - \mathbf{r}_i), \tag{2.63}$$

where $\delta(\mathbf{x})$ is the Dirac delta function. In thermodynamic equilibrium, ions obey Maxwell-Boltzmann statistics and since the energy of an ion under the influence of an electrostatic potential is $q_i\phi(\mathbf{x})$, the average density of mobile charges is

$$\rho_{\text{mob}}(\mathbf{x}) = \sum_j q_j n_j \exp\left(-\frac{q_j\phi(\mathbf{x})}{k_{\text{B}}T}\right), \tag{2.64}$$

where $n_j$ is the number density of ions in bulk solution far from any solute, while $k_{\text{B}}$ and $T$ denote the Boltzmann constant and temperature respectively, and the summation index $j$ iterates over ion species rather than over individual particles. Ions are only free to move inside the solvent, which is parameterised through an ion accessibility function, $\chi(\mathbf{x})$, whose value is 0 in regions occupied by solute and 1 elsewhere.

Substituting $\rho(\mathbf{x}) = \rho_{\text{fix}}(\mathbf{x}) + \rho_{\text{mob}}(\mathbf{x})$ into Equation 2.62 yields the Poisson-Boltzmann equation, which for neutral salts can be written as

$$-\nabla \cdot \big(\epsilon_0\epsilon(\mathbf{x})\nabla\phi(\mathbf{x})\big) + \bar{\kappa}^2(\mathbf{x}) \sinh\big(\phi(\mathbf{x})\big) = \rho_{\text{fix}}(\mathbf{x}), \tag{2.65}$$

where

$$\bar{\kappa}^2(\mathbf{x}) = \frac{e^2 I}{\epsilon_0 k_{\text{B}} T}\chi(\mathbf{x}) \tag{2.66}$$

represents ionic accessibility and screening effects. In the above, $e$ denotes elementary charge and

$$I = \frac{1}{2}\sum_j n_j \frac{q_j^2}{e^2} \tag{2.67}$$

is the ionic strength of the electrolyte.

Equation 2.65 is a nonlinear partial differential equation, which together with appropriate boundary conditions can be solved numerically through a multigrid finite difference method.[214,215] A solution is typically calculated on a cuboid domain comprising the bounding box of a solute protein plus a surrounding region filled with solvent. This domain is discretised into a regular grid and charges are assigned to grid points via a B-spline approximation of the density of atomic partial charges, $\rho_{\mathrm{fix}}(\mathbf{x})$.

Grid points belong to the solute occupied domain, if they are located within a distance of $R_i^{\mathrm{vdW}} + R^{\mathrm{sol}}$ from any of the solute atoms, and to the solvent occupied domain otherwise. Here $R_i^{\mathrm{vdW}}$ denotes the van der Waals radius of the $i$-th atom as specified in an MD force field and $R^{\mathrm{sol}} = 0.14\,\mathrm{nm}$ is the typical radius of a solvent molecule. Similarly, a grid point is considered ion accessible, if it is closer than $R_i^{\mathrm{vdW}} + R^{\mathrm{ion}}$ to any solute atom, where $R^{\mathrm{ion}}$ denotes the Born radius of the relevant ionic species. Both the permittivity and the ion accessibility are then smoothed by harmonic averaging over neighbouring grid points in order to reduce the sensitivity of the solution to the domain discretisation.[216]

Boundary conditions are typically set up through a focussing procedure, in which the Poisson-Boltzmann equation is first solved on a coarse grid using Dirichlet boundary conditions derived from Debye-Hückel theory.[217] Subsequently, a second solution is obtained on a fine grid focussed on a region of interest, where the potential at the boundary is set to the value obtained from the coarse grid calculation. All continuum electrostatics calculations in this thesis were carried out with the Adaptive Poisson-Boltzmann Solver (APBS).[218,219]

Give someone a program, you frustrate them for a day; teach them how to program, you frustrate them for a lifetime.

*(David Leinweber)*

# 3

# The Channel Annotation Package

Elucidating the molecular mechanisms underlying the gating of ion channels is an important objective for structural biology. As discussed in Section 1.1.2, recent advances in techniques such as cryo-electron microscopy are now generating an increasing wealth of new ion channel structures, including many captured in different conformational states. However, before this structural information can be compared to the detailed electrophysiological data that can often be obtained for many ion channels, the conductive status of any new structure needs to be known.

Conventionally, this has been achieved by determining the physical (i.e. steric) dimensions of the channel pore. A variety of approaches and biomolecular visualisation tools exist for determining the dimensions of pores, tunnels and cavities in proteins,[220] but by far the most common software used for mapping transmembrane ion channel pores are programs such as HOLE[221,222] and CAVER.[223,224] These calculate a radius profile along the conduction pathway and the structure is typically considered to be open and conductive if its narrowest constriction exceeds the radius of the hydrated ion species it conducts (Figure 3.1).

This size-based approach has served the ion channel community well for several decades, but an increasing body of evidence has demonstrated that it is unable to detect gates

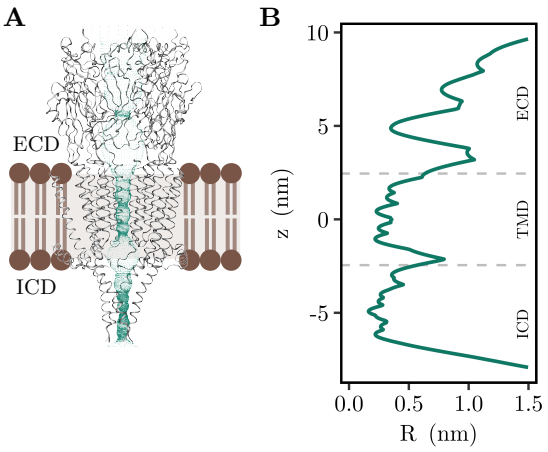



**Figure 3.1:** Physical dimensions of the ion conducting pore in a 5-$HT_3$ receptor. (A) HOLE profile of the entire PDB structure of the receptor (PDB ID: 4PIR). (B) Radius profile of the receptor pore. Data such as this is frequently used to assign a functional status to an ion channel structure. Figure adapted from Klesse et al. [26] under a creative commons (CC BY 4.0) licence.

that are not based on steric occlusion of the permeation pathway. As discussed in Section 1.2, hydrophobic gates can occur in channel pores to prevent ion flow and rely on the capillary evaporation of water that occurs when the lining of the pore is particularly hydrophobic.

In a previous study, Trick et al. [78] have shown how molecular dynamics (MD) simulations of water behaviour inside ion channel pores can be used to aid the functional annotation of different pentameric ligand-gated ion channels (the 5-$HT_3$ and glycine receptors) and demonstrated the existence of hydrophobic gates within these two receptors. In particular, it was established that pore hydration observed in simple equilibrium MD simulations can be used as a reliable proxy for ion permeability. This was validated by comparison to more computationally expensive and time-consuming approaches such as umbrella-sampling free energy calculations and computational electrophysiology simulations.[93]

This chapter now presents a versatile, new computational tool for use within the structural biology community that exploits the principle of pore hydration to rapidly predict the conductive state of new ion channel structures. Here the focus lies on the methodology, while the application of the new tool to relevant ion channel structures is discussed in Chapter 4. The remainder of this chapter is organised as follows. Section

3.1 provides an overview of how molecular dynamics can be employed for the functional annotation of ion channel structures. Subsequently, Section 3.2 discusses the implementation of this procedure as a robust software tool. Additional aspects of the underlying numerical algorithms and the details of the MD simulation protocol are explained in Section 3.3. The chapter concludes with a summary in Section 3.4. Results presented here have been published in Klesse et al. [26] and this chapter adapts figures and text from this publication under a creative commons (CC BY 4.0) licence.

## 3.1 Workflow for Functional Annotation

The workflow underlying this Channel Annotation Package (CHAP) is shown in Figure 3.2 and the required software is freely available (www.channotation.org). As an initial analysis step, CHAP can be run directly on the PDB structure of any channel or nanopore to generate both radius and hydrophobicity profiles of the putative ion conduction pathway. These initial steps are easily performed on an average desktop computer and do not require MD simulation of the structure. However, if these profiles suggest the channel may contain a hydrophobic gate or any other interesting features then a more in-depth analysis based on equilibrium MD simulation of the structure can be performed.

The example illustrated in Figure 3.2 is a crystal structure of a mammalian 5-$HT_3$ receptor (PDB ID: 4PIR) in which the existence of a hydrophobic gate has been demonstrated by previous studies.[78,225,226] Initial structural analysis with CHAP (i.e. without simulation) yields plots of both radius and hydrophobicity for the entire channel pore. The simple radius plot shows the size of the transmembrane section of the pore to be roughly constant at $\sim$0.3 nm radius. However, the accompanying hydrophobicity plot also reveals this section to be somewhat hydrophobic, especially in the region of the 9′ ring of side chains.

To test for the presence of a hydrophobic gate within this region, the structure was inserted in a lipid bilayer system ready for equilibrium MD simulations using GROMACS.

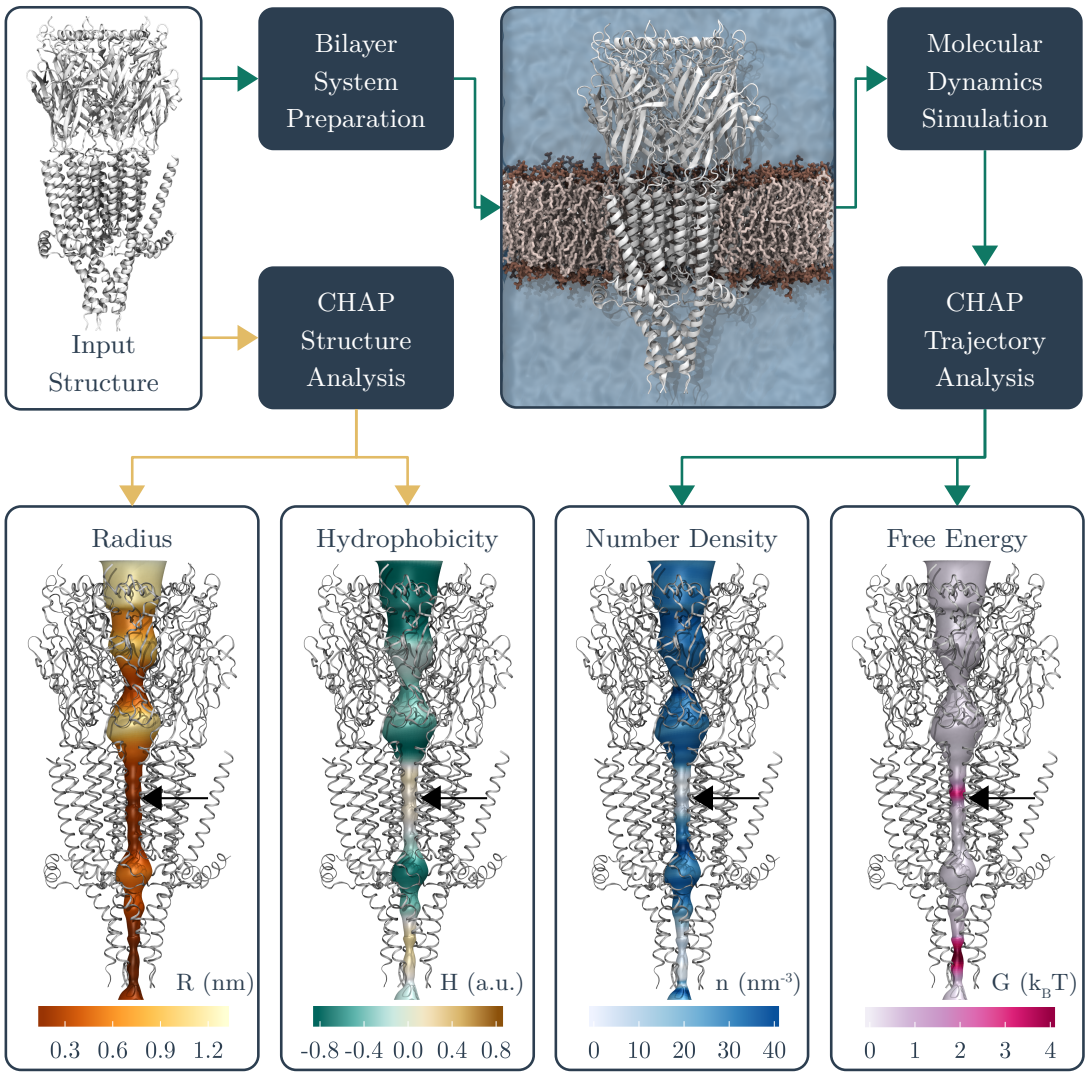



**Figure 3.2:** Overview of workflow for functional annotation of ion channels with CHAP. Given a simple protein structure as input, CHAP can calculate pore radius and hydrophobicity profiles. Additionally, the protein can be simulated in its native lipid bilayer environment in order to assess the behaviour of water inside the channel pore. From the resulting simulation trajectory data, CHAP can derive water density and free energy profiles that aid the identification of hydrophobic gates. Yellow arrows indicate steps of the pipeline that can be carried out on structural data alone while green arrows indicate those steps that are based on subsequent molecular dynamics simulation. The black arrows indicate the location of the 9' ring of leucine residues in the middle of the hydrophobic gate in the transmembrane domain of the channel. There is a further free energy barrier within the intracellular domain, which coincides with a narrowing of the radius of the pore. In an open structure of the 5-$HT_3$ receptor, ions are thought to permeate lateral portals within the intracellular domain, thus bypassing this constriction. Figure adapted from Klesse et al. [26] under a creative commons (CC BY 4.0) licence.

Briefly, the starting structure was embedded in a DOPC bilayer using a serial multi-scale procedure[227] and the resulting protein-bilayer system was then simulated using the CHARMM36m all atom force[228] field and the CHARMM-modified mTIP3P water model.[118] The C$\alpha$ atoms of all protein residues were harmonically restrained at their positions in the experimentally determined structure (see the Section 3.3.1 for details of the simulation parameters). These restraints prevent any gating transitions or other large-scale conformational transitions, whilst still permitting e.g. side chain reorientation. This ensures that any functional annotation produced by CHAP is as relevant as possible to the input structure itself, whilst simultaneously avoiding an unphysically rigid pore surface.

As shown in Figure 3.2, this allows calculation of the water density within the pore together with a free energy profile for hydration of the pore. The time-dependent behaviour of water during this particular simulation is also displayed in Figure 3.3. This shows that although the pore is fully hydrated in the initial configuration of the simulation system, water is expelled almost immediately (within <0.5 ns), leaving a 1 nm long region of the conduction pathway fully dehydrated. Consistent with our previous studies,[41] this rapid de-wetting was also observed within the first few nanoseconds in three independent MD simulations, each of which was run for a total of 100 ns. The rapid (<10 ns) nature of pore dewetting indicates that this simulation time is likely to be sufficient.

As discussed above, it is important to note that CHAP is designed to annotate specific experimentally determined structural states, i.e. it is not intended to simulate major conformational rearrangements. Thus relatively short simulation times are likely to be sufficient. Using this approach and the profiles calculated by CHAP, it becomes immediately apparent that the transmembrane pore of this particular conformation of the 5-$HT_3$ receptor contains a hydrophobic gate (Figure 3.2). This would not have been revealed if only a radius profile such as that shown in Figure 3.1 had been considered.

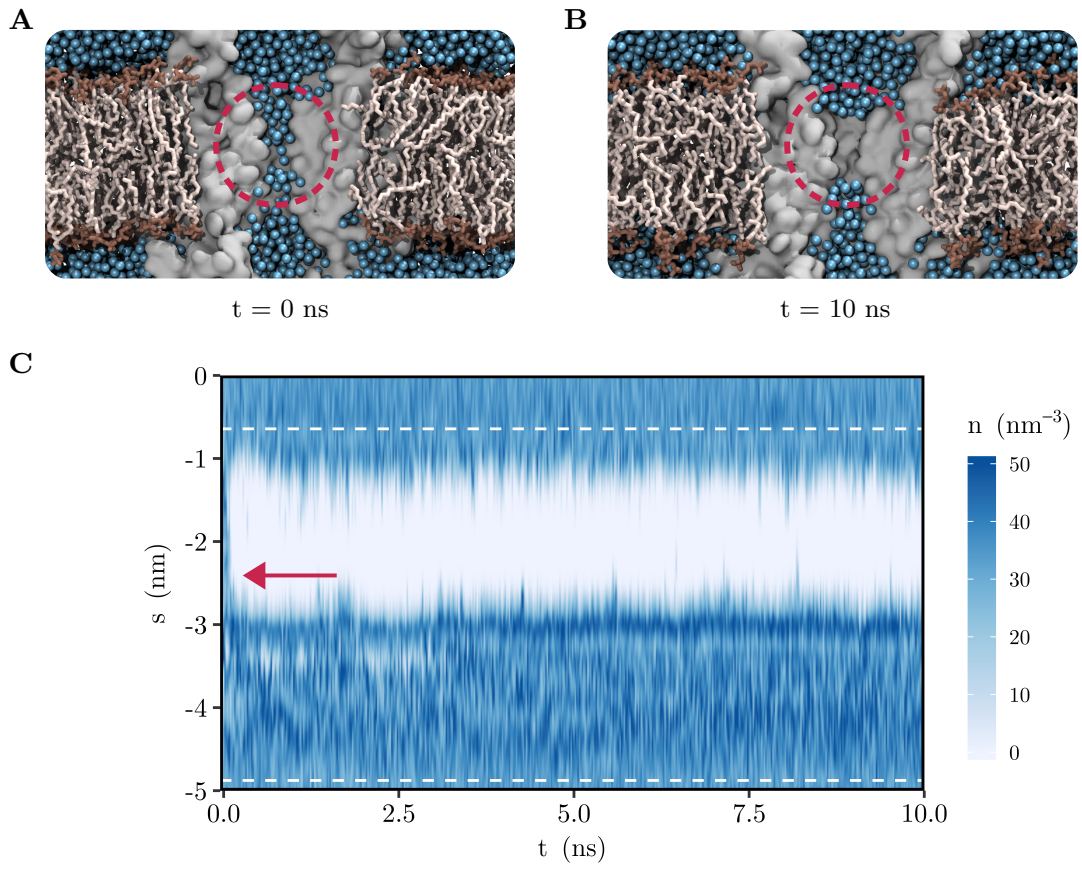



**Figure 3.3:** Hydrophobic gates can be identified through equilibrium MD simulation. (A) Cross-section of the starting structure for MD simulation of a 5-$HT_3$ receptor (PDB ID: 4PIR) embedded in a DOPC lipid bilayer. The protein is represented by a grey surface with two of five subunits omitted for visual clarity. Only the transmembrane (TM) domain is shown. Water oxygens are represented by blue spheres and lipid molecules are shown in liquorice representation. Note that the entire channel pore is initially filled with water. (B) Snapshot from the MD trajectory clearly exhibiting a de-wetted region (highlighted with red circle) in the TM domain of the channel. Note that the channel pore is not physically occluded. (C) Water density profile in the receptor pore over time. An approximately 2 nm long region of the TM domain indicated by a red arrow de-wets at the beginning of the simulation and remains devoid of water throughout. Figure adapted from Klesse et al. [26] under a creative commons (CC BY 4.0) licence.

## 3.2 Implementation of CHAP

Thus although tools that calculate pore radius alone have proven exceptionally useful it is clear that additional information is required if more accurate predictions are to be made. However, a simple extension of the popular HOLE program with added functionality was

not practical for several reasons. In particular, HOLE was originally developed for the analysis of static protein structures and is unable to read common trajectory data formats or process time-dependent structures similar to those generated by MD simulation. A new software suite for channel annotation was therefore implemented as an independent tool: CHAP is written in C++ and is based on the trajectory analysis framework of the popular MD simulation software GROMACS.[98]

The algorithm used by CHAP to calculate these different channel profiles is outlined in Figure 3.4. Starting from an input structure (either an experimentally derived PDB file or a single frame from an MD trajectory), the first step is to define the physical dimensions of the pore. A probe-based method is employed, similar to that used by HOLE, resulting in a sequence of discrete probe positions and associated pore radii. A continuous spatial curve is then determined from these probe positions by means of B-spline interpolation. Subsequently, a continuous radius profile is calculated by interpolating the probe radii along this curve (Figure 3.5).

This interpolating spline curve represents the centre line of the pore and defines a coordinate system that forms the basis for all subsequent calculations. In this channel coordinate system, the position of a particle (e.g. a water molecule or a protein side chain) is represented by its distance to the closest point on the centre line and the distance along the centre line between this point and a reference point (Figure 3.8A). By convention, the reference point is the projection of the centre of geometry of the overall channel protein onto the centre line.

Next, a hydrophobicity profile is determined for the residues that line the pore. By converting residue positions to channel coordinates, CHAP identifies all residues within a cutoff distance from the pore surface and then selects only those whose side chains point inwards towards the pore. The resulting discrete values for hydrophobicity at the positions of the individual residues are then converted into a continuous hydrophobicity profile along the centre line by means of kernel smoothing.[229]

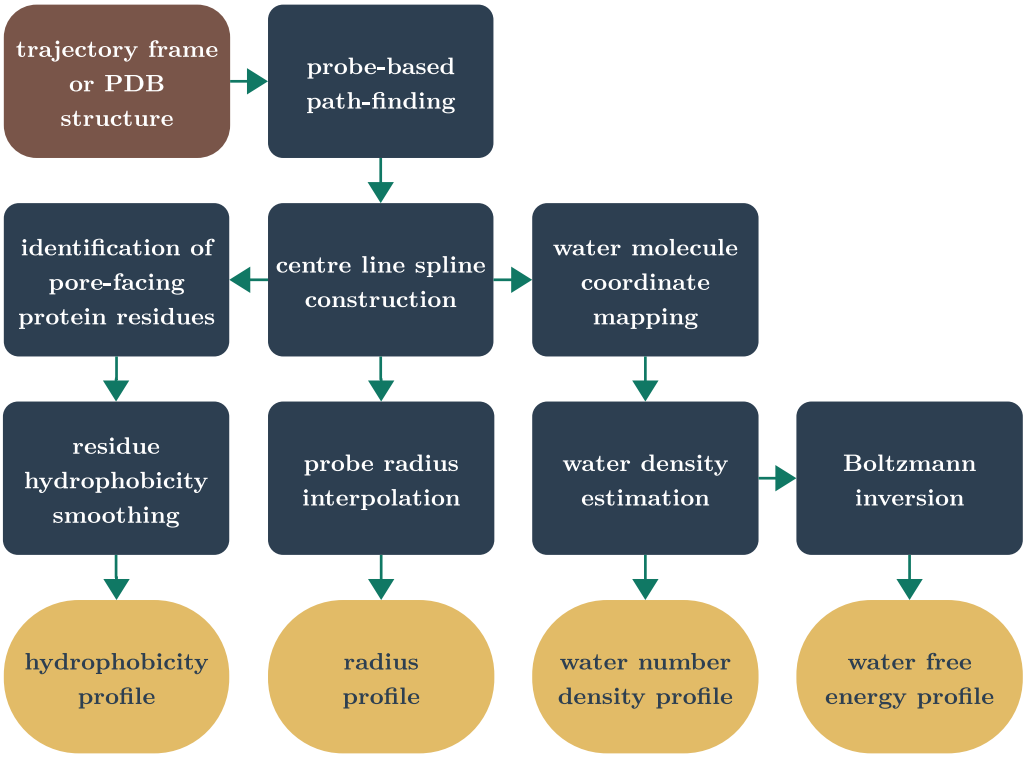



**Figure 3.4:** Workflow diagram of data processing in CHAP. For an individual PDB structure this procedure is carried out once. For trajectory data, the above procedure is carried out individually for each frame, i.e. for each structural configuration at a sampled time point. The resulting hydrophobicity, radius, density, and free energy profiles are averaged over all time steps and confidence bands are estimated from the standard deviation over time. Figure adapted from Klesse et al. [26] under a creative commons (CC BY 4.0) licence.

If the input structure contains water molecules, e.g. from an MD simulation, then CHAP will also estimate the solvent density along the centre line. To achieve this, water molecule positions are first transformed into channel coordinates, and then a one-dimensional probability density is calculated by kernel density estimation.[230] Together with the radius profile, this allows the calculation of the number density of water molecules throughout the channel structure. From these density values the related water free energy profile can then be estimated using Boltzmann inversion.

When the input to CHAP is an MD trajectory, the above procedure is repeated independently for each frame. CHAP will then calculate summary statistics (mean, standard

deviation, minimum, and maximum) over time for each point in the channel profiles. The following sections now describe these different analysis steps in more detail.

### 3.2.1 Pore Geometry and Centre Line

The path-finding algorithm used by CHAP is based on a hard sphere representation of the protein, in which each atom is associated with a sphere of radius $r_i$ centred at the position of the atom, $\mathbf{Q}_i$. Figure 3.5A illustrates how a pathway is found by squeezing a spherical probe through this collection of van der Waals spheres. Starting from an initial position inside the channel pore, $\mathbf{P}_0$, the probe is moved in steps of $\Delta v$ along the direction of a unit vector $\mathbf{e}_v$ pointing in the overall direction of the permeation pathway:

$$\tilde{\mathbf{P}}_i = \mathbf{P}_{i-1} + \Delta v \mathbf{e}_v. \tag{3.1}$$

The resulting point is then used to initialise a numerical optimisation procedure seeking to maximise the radius of the probe,

$$R(\mathbf{P}) = \min_{j \in [1,N]} ||\mathbf{Q}_j - \mathbf{P}|| - r_j, \tag{3.2}$$

without overlapping the van der Waals sphere of any protein atom. This optimisation procedure is carried out in the plane supported by $\tilde{\mathbf{P}}_i$ and spanned by the unit vectors $\mathbf{e}_u$ and $\mathbf{e}_w$ which are chosen to be perpendicular to both $\mathbf{e}_v$ and one another. The optimal probe position,

$$\mathbf{P}_i = \arg\max_{u,w} R(\tilde{\mathbf{P}} + u\mathbf{e}_u + w\mathbf{e}_w), \tag{3.3}$$

is accepted as a point on the pore centre line and a further step in the direction of $\mathbf{e}_v$ is carried out. This procedure is repeated until the probe radius exceeds a threshold, at which point it is assumed that the probe has left the pore. Subsequently, the iteration

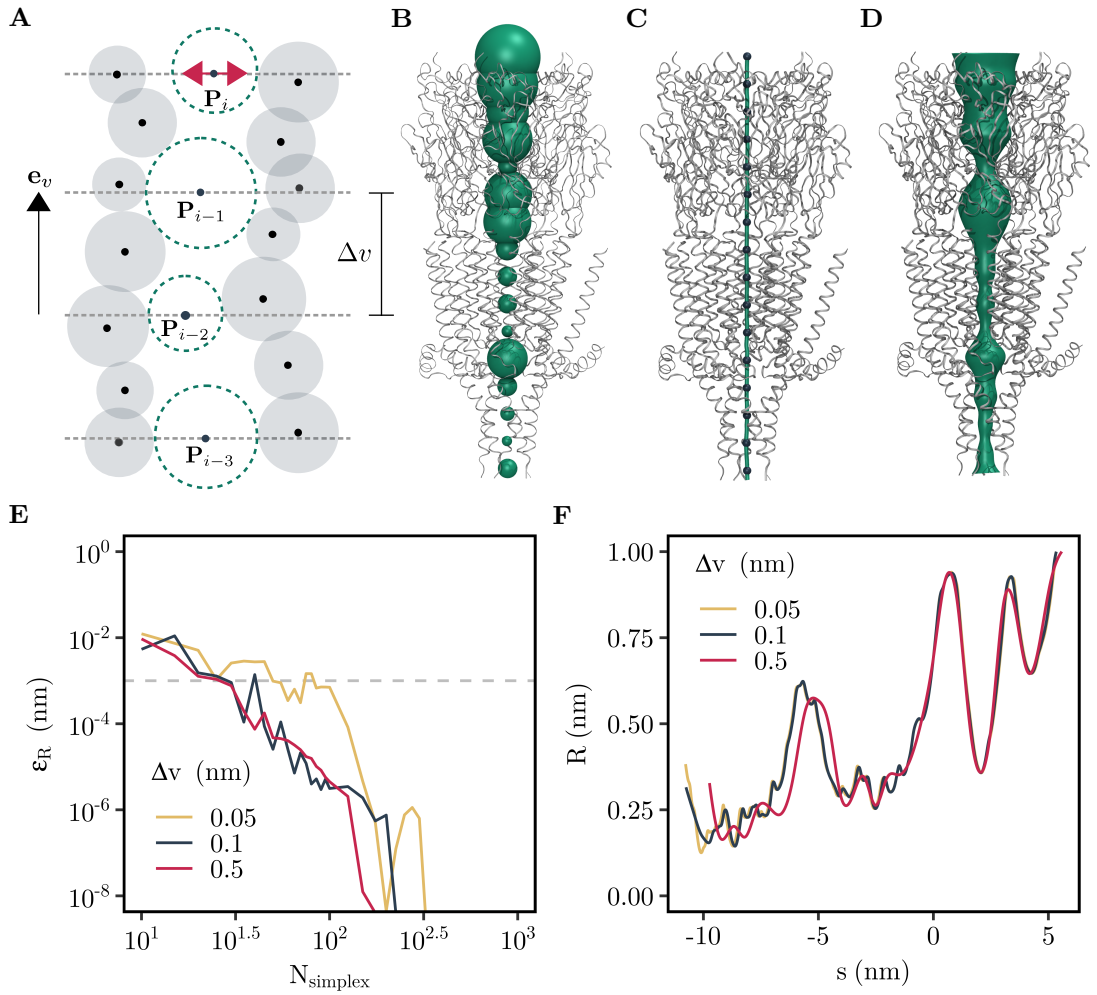



**Figure 3.5:** Pore geometry calculations in CHAP. (A) Schematic representation of probe-based pathway finding. The position of a spherical probe is optimised in subsequent parallel planes to maximise its radius without overlapping the van der Waals sphere of any ion channel atom. (B) Configuration space view of probe positions (green spheres) in the permeation pathway of a 5-$HT_3$ receptor structure (PDB ID: 4PIR). For visual clarity a large probe step of $v = 1$ nm was used here. (C) The probe centre positions (blue spheres) are interpolated with cubic B-splines to yield a continuous spatial curve (green tube) representing the pore centre line. (D) A surface representation of the channel pore is generated by extruding a circular cross-section along the centre line spline curve. (E) Convergence of the Nelder-Mead algorithm carried out to optimise the probe position in each parallel plane. The radius error falls below a threshold of 0.001 nm within 100 optimisation steps irrespective of the value of the probe step. The reference radius profile is calculated with a very large number of 2000 Nelder-Mead iterations. (F) Influence of probe step on radius profile resolution. While a probe step of 0.5 nm only captures the general feature of the radius profile, decreasing the probe step from 0.1 nm to 0.05 nm leads only to minimal changes. Figure adapted from Klesse et al. [26] under a creative commons (CC BY 4.0) licence.

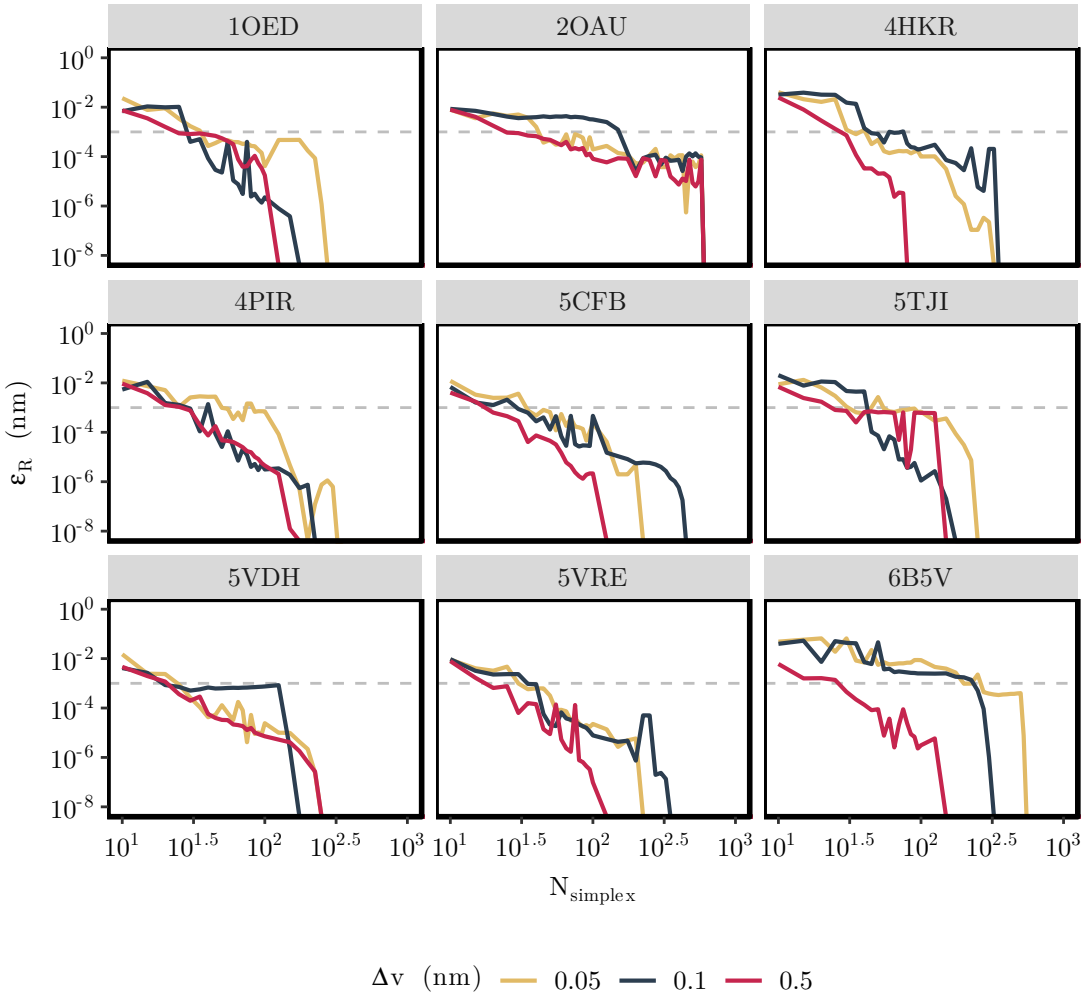



**Figure 3.6:** Convergence of Nelder-Mead optimisation for nine different channel structures. The radius error typically falls below a convergence threshold of 0.001 nm within ∼100 iterations of the simplex optimisation algorithm, irrespective of the step size, $\Delta v$, used in the path finding procedure. Figure adapted from Klesse et al. [26] under a creative commons (CC BY 4.0) licence.

is carried out a second time moving in the direction of $-\mathbf{e}_v$ and starting from point $\mathbf{P}_1$.

In the original algorithm used by HOLE the optimisation of the probe position was performed by means of simulated annealing.[222] However, being a global optimisation procedure, simulated annealing is prone to allowing the probe to jump out of the protein pore. The downhill simplex algorithm[231] was therefore implemented in CHAP alongside the simulated annealing procedure and was found to be generally more robust.

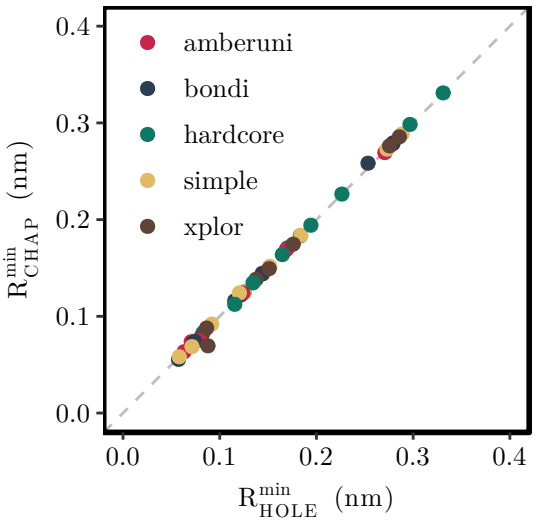



**Figure 3.7:** Comparison of the radius at the narrowest constriction between CHAP and HOLE. Calculated for nine representative ion channel structures using the five different van der Waals radius datasets available in HOLE. In all cases the radii agree to within 0.01 nm. Figure adapted from Klesse et al. [26] under a creative commons (CC BY 4.0) licence.

This procedure has three free parameters: the probe step size, $\Delta v$, the number of Nelder-Mead iterations, $N_{\text{simplex}}$, and the initial probe position, $P_0$. For a sufficiently symmetric protein with a central pore, $P_0$ can simply be set to the centre of geometry of the overall channel protein. The step size needs to be small enough to resolve all structural features of interest. As illustrated in Figure 3.5F, a probe step of 0.1 nm (the van der Waals radius of a hydrogen atom) is sufficient and the radius profile does not change significantly when the probe step is decreased further. Figure 3.5E shows the dependence of the error in the radius profile on the number of simplex iterations and reveals that 100 steps are sufficient to converge the profile to within a tolerance of ∼0.001 nm even for a very small probe step. This is the case not just for this particular 5-$HT_3$ receptor structure, but also for a wide range of other ion channel proteins (Figure 3.6).

The points $\mathbf{P}_i$ on the centre line are then connected into a continuous spatial curve $\mathbf{S}(s)$ by means of B-spline interpolation as detailed in Section 3.3.2 so that the curve passes exactly through all centre line points: $\mathbf{S}(s_i) = \mathbf{P}_i$. A corresponding continuous radius profile is then found by interpolating the probe radii along the centre line curve. The main advantage of a continuous centre line curve over individual points is that

the curve can be used to define a simple one-dimensional reaction coordinate for the motion of particles through the channel pore, as is detailed in the following section. Additionally, it can be used to generate a pore surface representation through the path extrusion technique as shown in Figure 3.5D. Importantly, for radially symmetrical pores the radius profiles calculated using CHAP are almost identical to those calculated using HOLE (Figure 3.7). However, it should be noted that the $s$-coordinate calculated by CHAP follows the pathway of the pore, and so for non-linear pathways this may differ from the $z$-coordinate calculated by HOLE.

An example of an ion channel, where this behaviour is useful, is the ganglionic nicotinic receptor. In a recent study[232] of a heteromeric structure of this protein, CHAP was used to determine the ion conduction pathway through five lateral portals located on the intracellular domain of the receptor. Since the orientation of these portals is almost perpendicular to the transmembrane segment of the channel pore, an algorithm based on a global Cartesian coordinate system would not have been appropriate in this case.

### 3.2.2 Water Density and Free Energy

It has previously been shown that water density can be used to estimate a free energy profile for pore hydration and how this acts as proxy for measurement of ion permeation.[78] To achieve this within CHAP, the centre line spline is used to define a curvilinear coordinate system, in which the position of each particle is described by its distance from the spline curve, $\rho$, the arc length parameter value corresponding to the closest point on the spline curve, $s$, and the angle, $\phi$, enclosed between the vector connecting this point and the particle position and the normal vector of the spline curve (Figure 3.8A). The resulting coordinate system can be viewed as a polar version of the Frenet-Serret frame, providing a generalisation of the canonical cylindrical coordinate system, where the $z$-coordinate is replaced by the length along the arc of the centre line.

Unfortunately, there is no analytical expression relating the Cartesian coordinates of

a particle to the curvilinear centre line coordinates and this mapping has to be accomplished numerically by finding the minimal distance between the position of a particle and the spline curve. Given the position of the $j$-th particle in Cartesian coordinates, $\mathbf{Q}_j = (x_j, y_j, z_j)^\intercal$, the corresponding spline coordinates are found from

$$\rho_j = \min_s ||\mathbf{S}(s) - \mathbf{Q}_j|| \tag{3.4}$$

$$s_j = \arg\min_s ||\mathbf{S}(s) - \mathbf{Q}_j|| - \arg\min_s ||\mathbf{S}(s) - \mathbf{P}_0|| \tag{3.5}$$

$$\phi_j = \frac{(\mathbf{Q}_j - \mathbf{S}(s_j)) \cdot \mathbf{N}}{||\mathbf{Q}_j - \mathbf{S}(s_j)||\ ||\mathbf{N}||} \tag{3.6}$$

where the minimisation problem is solved using the method of Brent [233]. Note that in the definition of the $s$-coordinate the projection of the initial probe position is subtracted. This ensures that the offset of the $s$-coordinate is consistent across multiple trajectory frames by setting it to zero for the initial probe position (which typically corresponds to the centre of geometry of the protein).

The $s$-coordinate is used as a one-dimensional collective variable along which the water density and the corresponding free energy profile can be determined. The probability density $P(s, \rho, \phi)$ of finding a water molecule at the given centre line coordinates is given by the Boltzmann relation,

$$P(s, \rho, \phi) = \frac{1}{Z} \exp\big(-\beta G(s, \rho, \phi)\big), \tag{3.7}$$

where $G(s, \rho, \phi)$ is the Gibbs free energy, $Z$ is the partition function, and $\beta^{-1} = k_\mathrm{B}T$ with $k_\mathrm{B}$ being the Boltzmann constant and $T$ the simulation temperature. In order to obtain a one-dimensional free energy profile, i.e. $G(s, \rho, \phi) = G(s)$, the lateral degrees of freedom are integrated out according to

$$P(s) = \int_0^{2\pi} \int_0^{R(s)} P(s, \rho, \phi) \rho d\rho d\phi = \frac{\pi R(s)^2}{Z} \exp\big(-\beta G(s)\big), \tag{3.8}$$

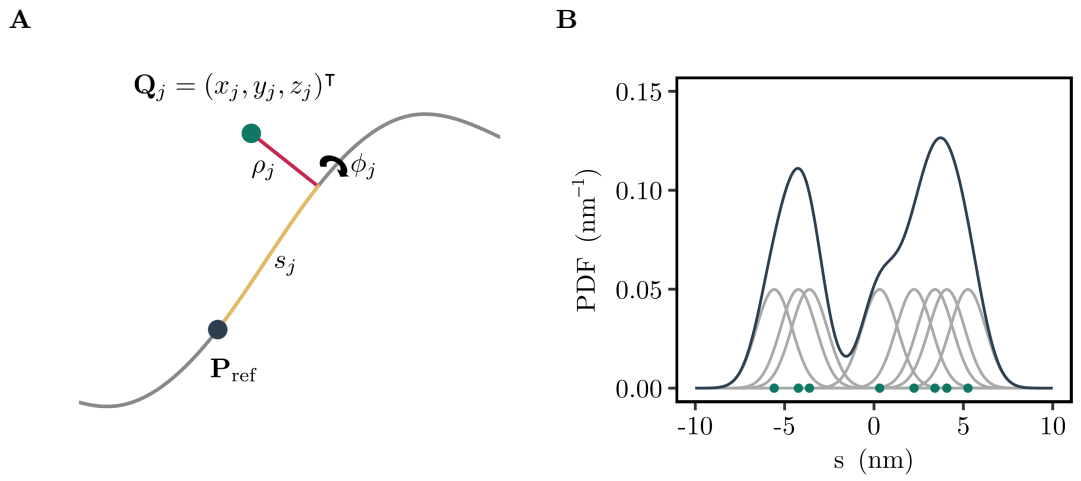



**Figure 3.8:** Determination of water density. (A) Cartesian coordinates of a water molecule (green) are converted into a curvilinear coordinate system by projecting its position onto the channel centre line (grey). The new coordinates are then expressed in terms of the distance along the centre line (yellow) between the projection base point and a reference point (dark blue), the distance from the water molecule to the centre line (red), and an angle around the centre line. (B) Conceptual representation of kernel density estimation: Each discrete water molecule position is associated with a Gaussian probability distribution and the sum of all Gaussians yields the probability density function of water along the channel centre line. Figure adapted from Klesse et al. [26] under a creative commons (CC BY 4.0) licence.

where it is assumed that the cross-sectional area of the pore is approximately circular with a local radius $R(s)$. Inverting the above relationship yields an expression for the free energy of the water molecule,

$$G(s) = -\frac{1}{\beta}\left(\ln \frac{P(s)}{\pi R(s)^2} + \ln Z\right), \tag{3.9}$$

into which the partition function enters only as an additive constant. This constant is determined by the convention that the free energy in the bulk outside the protein should be zero, i.e. $G(s) \to 0$ as $s \to \pm\infty$.

The probability density appearing in the above expression can be estimated directly

from the MD trajectory by means of kernel density estimation:

$$P_t(s) = \frac{1}{N} \sum_{i=1}^{N} K_h\big(s, s_i(t)\big) \tag{3.10}$$

Here $s_i(t)$ is the position of the $i$-th water molecule at time $t$, $N$ is the overall number of water molecules, and $P_t(s)$ denotes the probability density estimated at time $t$. A Gaussian kernel of the form

$$K_h(s, s_i) = \frac{1}{\sqrt{2\pi}h} \exp\left(-\frac{s - s_i}{2h^2}\right) \tag{3.11}$$

is employed here. A conceptual representation of kernel density estimation is shown in Figure 3.8B.

The bandwidth $h$ is a free parameter in kernel density estimation that fulfils a function similar to the bin width in histograms. Care must be taken when selecting a value for this parameter as the resulting density profile is sensitive to it. A value for this parameter of 0.24 nm has been suggested for identifying liquid interfaces in three dimensions,[234] whereas a value of 0.1 nm has been used in a study of pore de-wetting in the bacterial GLIC channel.[72] It is also possible to estimate an appropriate bandwidth directly from the data.[230] This method minimises the asymptotic mean integrated square error (AMISE) of the density estimate.

Figure 3.9 uses the example of the 5-$HT_3$ receptor to illustrate how the water density and free energy profiles depend on the bandwidth parameter. For very small values of $h$, the water density estimate is very noisy and in low density regions it drops down to zero, so that the corresponding free energy at these points is singular. Conversely, if a very large bandwidth is chosen, the estimate of $P(s)$ is near uniform throughout the pore and the water density estimate exhibits over-densities at points where the pore radius is particularly small. In this case the free energy profile exhibits no clear barriers or wells.

The AMISE-optimal bandwidth, which in this case is 0.34 nm, finds a balance between

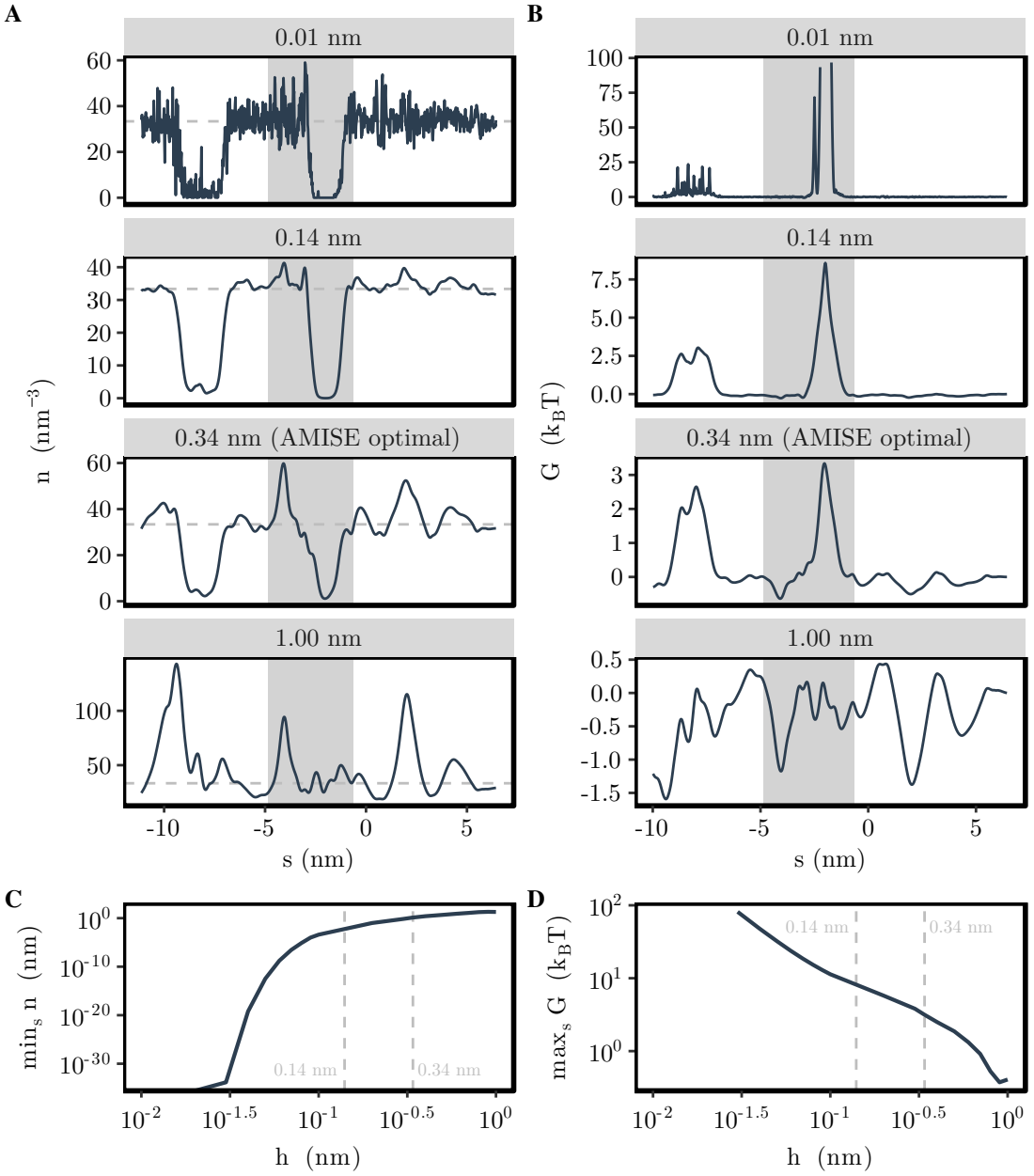



**Figure 3.9:** Influence of bandwidth parameter. (A) Water number density along the permeation pathway of the 5-HT$_3$ receptor for various values of the kernel bandwidth. A value of 0.14 nm corresponds to the approximate size of a water molecule and a value of 0.34 nm corresponds to the statistically optimal bandwidth. (B) Corresponding free energy profiles calculated via Boltzmann inversion of the density profiles. The grey shaded area in (A) and (B) represents the transmembrane domain. (C) Minimum of number density profile in dependence of bandwidth. (D) Maximum of free energy profile in dependence of bandwidth. Figure adapted from Klesse et al. [26] under a creative commons (CC BY 4.0) licence.

these two extremes. The overall density profile is free of noise and the lowered water density in the region of the hydrophobic gate is readily apparent. However, there are still some regions where the water density is unphysically large. By comparison, a profile estimated using a bandwidth of 0.14 nm (the approximate radius of a water molecule) yields a water density that never much exceeds the experimental bulk density of water while still showing little sign of noise. Using a bandwidth fixed at the size of a water molecule also permits a more direct comparison between different channel structures, while the AMISE-optimal bandwidth may differ across channel structures of different overall pore volume. It is therefore recommended to use a default value of 0.14 nm.

### 3.2.3 Hydrophobicity Profiles

In addition to calculating the channel radius and water density profiles inside the pore, it is also necessary to quantify the hydrophobicity along the conduction pathway. For cylindrical nanopores, it has been found that the free energy difference between hydrated and dehydrated pore states is given by

$$\Delta\Omega = \Omega_{\text{vapour}} - \Omega_{\text{liquid}} = 2\pi R\gamma_{\text{lv}}(R + L\cos(\theta_c)), \tag{3.12}$$

where $R$ and $L$ are the radius and length of the nanopore, $\gamma_{\text{lv}}$ is the liquid-vapour surface tension of water, and $\theta_c$ is the contact angle from the Young equation (see Section 1.2.4 for a detailed discussion). Determining a hydrophobicity profile for a channel therefore requires knowledge about the amino acid residues facing a given section of the pore and their associated contact angle. The individual hydrophobicity values at the positions of the amino acids can then be combined into a continuous hydrophobicity profile by means of kernel smoothing.

CHAP can automatically identify pore-facing residues, i.e. residues exposing their amino acid side chains to the channel pore (Figure 3.10A), via a two-step procedure.

First, all residues whose centre of geometry lies within a given cutoff, $R_{\text{thres}}$, from the pore surface are determined. To accomplish this, the position of the centre of geometry of each residue is mapped from Cartesian coordinates to centre line spline coordinates so that each residue is associated with a distance from the centre line, $\rho_i^{\text{cog}}$, and a distance along the centre line, $s_i^{\text{cog}}$, where the index runs over all protein residues. If

$$\rho_i^{\text{cog}} > R(s_i^{\text{cog}}) + R_{\text{thres}}, \tag{3.13}$$

the $i$-th residue will be considered pore-lining, where $R(s_i^{\text{cog}})$ is the pore radius at the position of the residue. A typical threshold value for an $\alpha$-helical pore is $\sim$0.75 nm.

Second, for each pore-lining residue, the position of the $\alpha$-carbon is also mapped to pathway coordinates $\rho_i^{\alpha}$ and $s_i^{\alpha}$. If the distance from the $\alpha$-carbon to the pore centre line is larger than the distance between the centre of geometry of the overall residue and the pore centre line, i.e. if $\rho_i^{\alpha} > \rho_i^{\text{cog}}$ then the $i$-th residue will be considered pore-facing. The property of being pore-lining and pore-facing is encoded in two binary indicator variables, whose values are written to the temperature factor and occupancy fields of a PDB file of the input structure to allow easy visualisation of these residues. If CHAP is run with a trajectory as input, the above analysis is performed repeatedly (once for each frame) and the values written to the PDB file correspond to the time average of the indicator variables. Figure 3.10A illustrates the distinction between pore-lining and pore-facing residues in the 5-$HT_3$ receptor.

To the best of our knowledge no experimental data quantifying the contact angle at the interface between water and specific amino acids are available. However, various amino acid hydrophobicity scales have been proposed in the literature. An early hydropathy scale based on a collection of experimental observations has been suggested by Kyte and Doolittle [235], whilst another has been based on measurements of the stability of $\alpha$-helices.[236] Other scales have used the free energy of partitioning of oligopeptides between

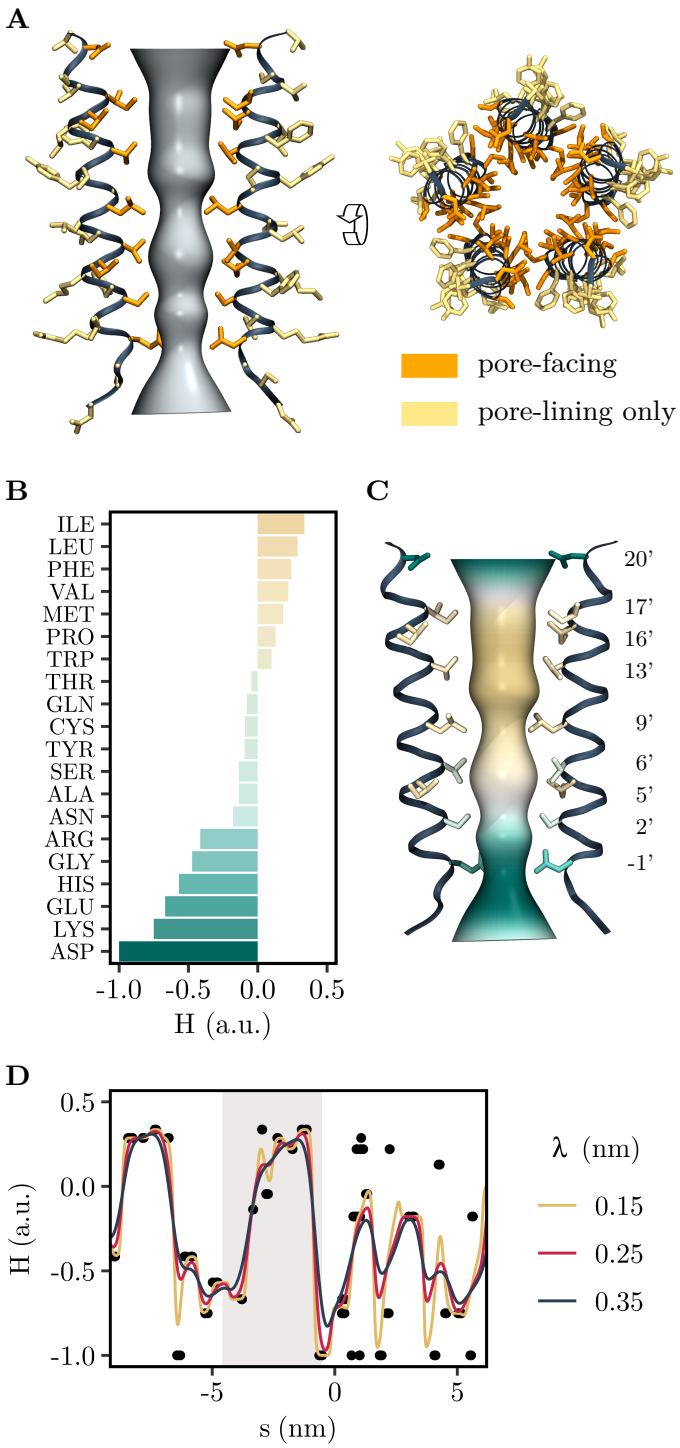



**Figure 3.10:** Hydrophobicity profile of pore-lining residues. (A) CHAP automatically subdivides the pore-lining residues into two categories: those whose side chains directly face the pore to influence the permeation pathway (i.e. pore-facing residues), and those whose side chains are close to, but point away from the pore (i.e. pore-lining only). In this example of the M2 helix bundle of the 5-$HT_3$ receptor TMD, the 'pore-facing' residues are shown in orange, whilst the 'pore-lining only' residues are shown in yellow. (B) Amino acid hydrophobicities according to the scale proposed by Wimley and White. In CHAP the original hydrophobicities are rescaled into the interval from -1 (very hydrophilic) to +1 (very hydrophobic) to facilitate an easy comparison between different hydrophobicity scales. (C) Permeation pathway coloured according to the Wimley-White hydrophobicity of the pore-facing residues. (D) Hydrophobicity profile of the entire 5-$HT_3$ receptor pore. Black circles represent the position and hydrophobicity of the pore-facing residues (averaged over all five subunits). Lines represent the kernel-smoothed continuous hydrophobicity profile underlying the surface colouring in panel (C). Different values of the smoothing span, $\lambda$, influence the exact magnitude of hydrophobicity fluctuations, but the overall distinction between hydrophobic and hydrophilic regions along of the pore remains unaffected. The shaded area corresponds to the transmembrane region shown in panels (A) and (C). Figure adapted from Klesse et al. [26] under a creative commons (CC BY 4.0) licence.

water and a lipid environment,[237,238] while some report free energies of transferring amino acids from water into a lipid bilayer.[239] A biological scale has also been proposed based on the recognition of $\alpha$-helices by the endoplasmic reticulum translocon,[240] whilst another is derived from MD simulations of a water nanodroplet on a planar layer of amino acids.[241]

For comparability, the hydrophobicity scales listed above have been shifted and scaled to the interval ranging from -1 for the most hydrophilic to +1 for the most hydrophobic residues while preserving their natural zero. Note that this contrasts with the physical definition of hydrophobicity in terms of $\cos\theta_c$ and instead follows the biochemical convention of associating a larger value with a more hydrophobic residue. While the exact hydrophobicity value assigned to each amino acid differs to some degree across scales, Rao et al. [103] found no marked difference in their ability to predict pore de-wetting in a recent analysis of nearly 200 ion channel structures. By default, CHAP employs the Wimley and White scale,[237,238] but users are free to select an alternative scale that may be more appropriate for their structure.

CHAP uses the normalised hydrophobicity scales to associate a hydrophobicity value, $H_i$, with each residue position $s_i^{\mathrm{cog}}$. A continuous hydrophobicity profile is then calculated from these discrete hydrophobicity values using a Nadaraya-Watson kernel-weighted average. Mathematically this is expressed as

$$H(s) = \frac{\sum_i K_\lambda(s, s_i) H_i}{\sum_i K_\lambda(s, s_i)}, \tag{3.14}$$

where

$$K_\lambda(s, s_i) = \exp\left(-\frac{|s - s_i|^2}{2\lambda^2}\right) \tag{3.15}$$

is a Gaussian kernel. The smoothing span $\lambda$ is chosen to reflect the finite extension of an amino acid residue. Since the pitch of an $\alpha$-helix is approximately 0.15 nm per residue

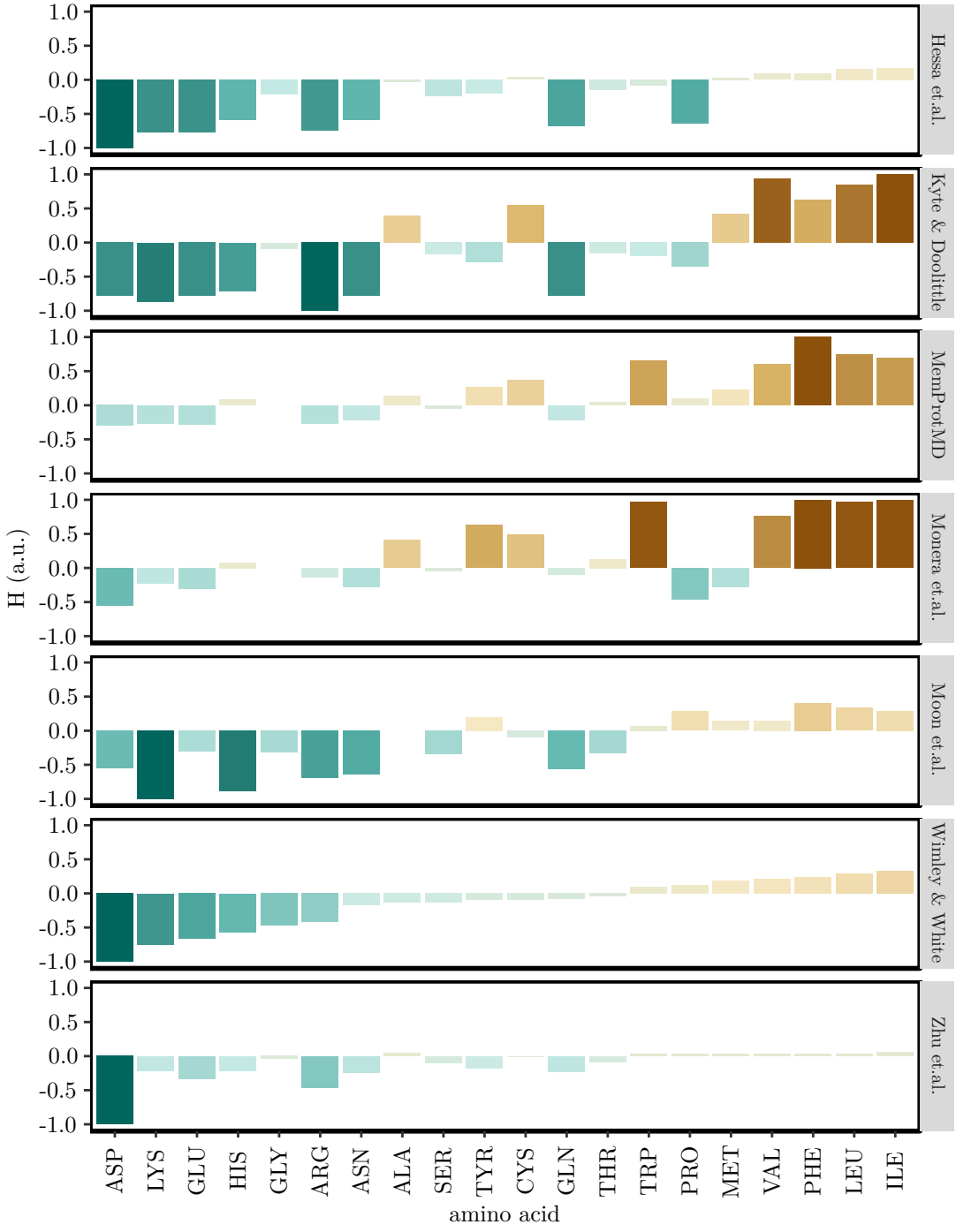



**Figure 3.11:** Comparison of amino acid hydrophobicity scales available in CHAP. The amino acids are listed in order of increasing hydrophobicity according to the Wimley-White scale, which is the default scale used by CHAP. Figure adapted from Klesse et al. [26] under a creative commons (CC BY 4.0) licence.

and on average every third residue is pore-facing, a value of ∼0.225 nm (for a width of the Gaussian kernel of ∼0.45 nm) can be considered reasonable. Note that far away from any residues (e.g. in the bulk water region outside the channel pore) $H(s)$ will decay exponentially to zero.

In Figure 3.10D this is illustrated on the 5-$HT_3$ receptor using the Wimley and White hydrophobicity scale. As can be seen, there is a considerable increase in hydrophobicity in the region of the transmembrane domain containing the hydrophobic gate. This section of the profile is shown in Figure 3.10C mapped onto a 3D representation of the pore together with the M2 helices of the 5-$HT_3$ receptor, illustrating how the hydrophobicity of the pore-facing residues influences the hydrophobicity profile. Alternative hydrophobicity scales available in CHAP are compared in Figure 3.11.

## 3.3 Further Implementation and Simulation Details

### 3.3.1 Molecular Dynamics Simulation Protocol

Equilibrium MD simulations of the 5-$HT_3$ receptor were performed using GROMACS 2018. Starting structures were generated from the original crystal structure (PDB ID: 4PIR) with missing atoms being added by the WHAT IF tool.[242] This starting structure was embedded in a DOPC bilayer using an established serial multiscale protocol.[227] The resulting protein-bilayer system was converted back to an atomistic representation and solvated in 150 mmol $L^{-1}$ NaCl solution. The CHARMM36m all-atom force field[228] was used together with the CHARMM-modified mTIP3P model of water.[118]

Long-range electrostatic interactions are treated using the smoothed particle mesh Ewald (PME) method[151] employing a short range cutoff of 1 nm and a Fourier spacing of 0.12 nm. To permit subsequent determination of the Gibbs free energy from the water molecule distribution, the system is simulated in the isothermal-isobaric (NPT) ensemble at a temperature of 310 K and a pressure of 1 bar. A v-rescale thermostat[142]

with a coupling constant of 0.1 ps is used for temperature control and a pressure is maintained semi-isotropically using the method of Parrinello and Rahman [143] with a coupling constant of 1 ps.

The canonical leapfrog method with a time step of 2 fs is used to integrate three independent copies of the system for 100 ns with bonds constrained through the linear constraints solver algorithm LINCS.[139] Additionally, the C$\alpha$ atoms of all protein residues are placed under a harmonic restraint with a force constant of 1000 kJ mol$^{-1}$ nm$^{-2}$ to ensure that the conformation of the simulated channel does not deviate from the experimentally determined structure. A simulation length of 100 ns likely exceeds what is required, since the largest permitted protein motions are side chain reorientations, which typically occur on a time scale of $\sim$10 ns and most pore dewetting events appear to occur even more rapidly.

### 3.3.2 B-Spline Curves

As explained in Section 3.2.1, B-splines are used internally by CHAP to represent pathways through three-dimensional protein structures as continuous curves. This section documents the algorithms underlying these geometrical calculations. Formally, a spline curve of degree $p$ along some parameter $s$ can be written as the linear combination

$$\tilde{\mathbf{S}}_p(\tilde{s}) = \sum_{k=1}^{N} \mathbf{c}_k B_{k,p}(\tilde{s}), \tag{3.16}$$

where

$$B_{k,p}(\tilde{s}) = \frac{\tilde{s} - t_k}{t_{k+p} - t_k} B_{k,p-1}(\tilde{s}) + \frac{t_{k+p+1} - \tilde{s}}{t_{k+p+1} - t_{k+1}} B_{k+1,p-1}(\tilde{s}), \tag{3.17}$$

with $B_{k,0}(\tilde{s}) = 1$ if $t_k \leq \tilde{s} \leq t_{k+1}$ and $B_{k,0}(\tilde{s}) = 0$ otherwise are the recursively defined B-spline basis functions. Here $\{\mathbf{c}_k\}_{k=1}^{N}$ denotes the spline curve control points and $\{t_k\}_{k=1}^{N+p+1}$ is referred to as the knot vector. Note that a zero division in the above equations is treated as $0/0 = 0$.

In order to find the spline curve that interpolates the centre line points $\mathbf{P}_i$ from the pore finding algorithm, one demands

$$\mathbf{S}_p(\tilde{s}_i) = \mathbf{P}_i, \tag{3.18}$$

which requires a parameterisation of the discrete point set. Ideally, this should be the distance between the points along the arc of the spline curve, but unfortunately this distance is not known a priori. Instead, the commonly applied chord length approximation is used here, in which points are parameterised by their Euclidean distance according to:

$$\tilde{s}_{i+1} = \tilde{s}_i + ||\mathbf{P}_{i+1} - \mathbf{P}_i||. \tag{3.19}$$

Importantly, it has been shown that this choice of parameterisation guarantees full approximation order for spline curves up to the third degree.[243]

For a cubic spline curve $\mathbf{S}(\tilde{s})$, where the subscript $p = 3$ has been dropped for clarity, Equation 3.18 together with the knot vector

$$\{t_k\}_{k=1}^{N+p+1} = \{\tilde{s}_1, \tilde{s}_1, \tilde{s}_1, \tilde{s}_1, \tilde{s}_2, \tilde{s}_3, \ldots, \tilde{s}_{N-1}, \tilde{s}_N, \tilde{s}_N, \tilde{s}_N, \tilde{s}_N\} \tag{3.20}$$

and hermite boundary conditions then leads to a tridiagonal linear system, from which the control points can be determined efficiently through the Thomas algorithm.[244] The choice of cubic splines is motivated by the fact that cubic splines are twice continuously differentiable so that the curvature of the resulting spline curve is guaranteed to be continuous. At the same time, cubic splines are known to minimise overall curvature, preventing undue variation of the interpolant between the given centre line points.

Subsequently, the resulting spline curve is re-parameterised by arc length using a three-step procedure proposed by Wang et al. [245]. In the first step, the arc length in

each interval is computed according to

$$l_k = \int_{\tilde{s}_k}^{\tilde{s}_{k+1}} ||\tilde{\mathbf{S}}(\tilde{s})||\, d\tilde{s} = \int_{\tilde{s}_k}^{\tilde{s}_{k+1}} \sqrt{\left(\tilde{\mathbf{S}}'(\tilde{s})\Big|_x\right)^2 + \left(\tilde{\mathbf{S}}'(\tilde{s})\Big|_y\right)^2 + \left(\tilde{\mathbf{S}}'(\tilde{s})\Big|_z\right)^2}\, d\tilde{s} \tag{3.21}$$

from which the overall length of the spline curve between the two openings of the pore can be calculated as:

$$L = \sum_{k=1}^{N-1} l_k. \tag{3.22}$$

In Equation 3.21, the integral is solved numerically through sixth-order accurate Newton-Cotes quadrature (Boole's rule).

The second step comprises calculating the chord length parameter values $\tilde{s}_m$ corresponding to the equidistant arc length parameter values $s_m = m\Delta s = mL/M$ with $m \in [0, M]$. This is equivalent to finding the root of

$$m\Delta s - \left(\sum_{k=1}^{q} l_k + \int_{\tilde{s}_q}^{\tilde{s}_m} ||\tilde{\mathbf{S}}'(\tilde{s})||\, d\tilde{s}\right) = 0, \tag{3.23}$$

where $q \in [1, N]$ is the index of the spline interval for which

$$\sum_{k=0}^{q} l_k \leq m\Delta s < \sum_{k=0}^{q+1} l_k. \tag{3.24}$$

Here the root-finding problem posed by Equation 3.23 is solved through the algorithm of Alefeld et al. [246], which is known to be the asymptotically most efficient derivative-free root-finding method.

In the third and final step, the interpolation problem

$$\mathbf{S}(s_m) = \tilde{\mathbf{S}}(\tilde{s}_m) \tag{3.25}$$

is solved to obtain a new spline curve, $\mathbf{S}(\mathbf{s})$, that approximates the original curve, but is parameterised by arc length rather than chord length. This curve is then used to

describe the centre line of the channel pore as described in Section 3.2.1.

## 3.4 Summary and Conclusions

This chapter introduced a versatile computational tool for the functional annotation of ion channel structures: the Channel Annotation Package, CHAP. This tool implements a methodology for predicting the conductive state of ion channel structures that goes beyond a measurement of the pore radius to include profiles of pore hydrophobicity as well as of water density and free energy based on MD simulations of water behaviour within an experimentally determined channel structure. Crucially, this enables CHAP to identify hydrophobic gates that cannot be detected using methods based purely on measuring the physical dimensions of the pore alone.

Importantly, the analysis performed by CHAP is computationally inexpensive compared to more complex simulation techniques such as computational electrophysiology[93] and is therefore highly suitable for the high-throughput analysis of the rapidly growing number of ion channel structures. The utility of this approach is demonstrated in a recent study[103] employing CHAP to investigate the prevalence of hydrophobic gates in nearly 200 different ion channel structures.

CHAP has been made available to the scientific community as free and open-source software. Copies of the source code can be obtained from www.channotation.org and an associated Github repository[247] and include scripts for visualising results in the molecular graphics programmes VMD[248] and PyMOL.[249] In addition, CHAP has been adopted by the SBGrid consortium,[250] an international organisation providing software infrastructure for structural biology labs, and is available to its ∼350 member groups as a fully serviced installation.[251]

Since the manuscript introducing CHAP[26] was first released on bioR$\chi$iv in January 2019, the programme has been employed in several investigations of biological ion channels and pores. These early adopters include structural studies of the ganglionic nicotinic

receptor[232] and the bacterial lipid transport protein LetB,[252] as well as a computational study exploring the refinement of cryo-EM structures through MD simulation.[253] In addition, CHAP was recently used in a compuational study of synthetic nanopores composed of cyclic peptides,[254] indicating the utility of the tool for users beyond its original target audience of structural biologists.

Still round the corner there may
wait, a new road or a secret gate.

*(J. R. R. Tolkien)*

# 4

# Hydrophobic Gates in Ion Channels

The previous chapter introduced the Channel Annotation Package, CHAP, as a versatile tool for the functional annotation of ion channels, based on equilibrium MD simulations of water. In this chapter, CHAP will be applied to analyse the pore hydration behaviour of several recently published ion channel structures. The utility of this new tool will thus be demonstrated by identifying hydrophobic gates in two members of the pentameric ligand-gated ion channel (pLGIC) family.

As a first example, the 5-hydroxytryptamine type-3 (5-$HT_3$) receptor will be considered. A closed state X-ray structure[255] of this cation-selective channel was already used in Chapter 3 to illustrate the functional annotation procedure implemented by CHAP. Here, that analysis will be extended to more recent cryo-EM structures of the 5-$HT_3$ receptor, which include an open and a desensitised conformation of the channel[225] as well as an additional closed state structure.[256]

Another pLGIC, the anion-selective glycine receptor, will serve as a second example. The closed and desensitised conformations of this channel have been elucidated by both X-ray crystallography[257,258] and cryo-EM experiments.[259] In addition, Du et al. [259] determined a cryo-EM structure of the open conformation of the glycine receptor, although there continues to be uncertainty regarding the physiological significance of this

| PDB ID | organism | expression system | method | resolution | state |
|---|---|---|---|---|---|
| 4PIR[255] | *Mus musculus* | *Escherichia coli* | X-ray | 0.350 nm | closed |
| 6BE1[256] | *Mus musculus* | *Spodoptera frugiperda* | cryo-EM | 0.431 nm | closed |
| 6DG7[225] | *Mus musculus* | *Spodoptera frugiperda* | cryo-EM | 0.332 nm | desensitised |
| 6DG8[225] | *Mus musculus* | *Spodoptera frugiperda* | cryo-EM | 0.389 nm | open |

**Table 4.1:** Structures of the 5-$HT_3$ receptor discussed in this thesis. Four additional structures have been published by Polovinkin et al. [263], but are not considered here. Note that the functional state indicated in the final column reflects the results presented in this chapter. Hassaine et al. [255] were unable to assign a functional state to the 4PIR structure, while Basak et al. [256] identify the 6BE1 structure with the closed or resting state of the channel. Basak et al. [225] refer to the 6DG7 and 6DG8 structures as state 1 and state 2, which through MD simulation are identified as nonconducting and conducting conformations respectively. In addition, the 6DG7 structure is suggested to represent a desensitised state due to its conformational similarity with the desensitised conformations of other channels in the Cys-loop receptor family.

structure.[260–262]

Simulations of the 5-$HT_3$ and glycine receptors will also be used to probe the robustness of the simulation protocol introduced in Section 3.3.1. In particular, the influence of position restraints will be analysed systematically in order to ascertain that a harmonic restraining force does not cause artefacts such as the spurious dewetting of a channel pore. Additionally, the possibility of simulating a subset of a channel structure, e.g. its transmembrane domain or its pore-lining helices, and ignoring structural features far from the location of the channel gate, such as extra- or intracellular domains, will be explored. The results presented for the 5-$HT_3$ and glycine receptors are in part based on work published in Trick et al. [78].

After analysing these two pLGICs in detail, this chapter will also demonstrate the use of CHAP as a high throughput tool by examining ∼200 ion channel structures for potential hydrophobic gates. This analysis will provide an indication of how prevalent hydrophobic gates are in the ion channel structural proteome and under what conditions of pore hydrophobicity and radius hydrophobic gating is likely to arise. These results are based on data published in Rao et al. [103].

The remainder of this chapter begins with an overview of the computational methods used in Section 4.1. Following this, simulations of the 5-$HT_3$ and glycine receptors are discussed in Sections 4.2.1 and 4.2.2 respectively, while Section 4.2.3 describes the results of the high throughput analysis of ion channels. Results are discussed and compared to other computational efforts for functionally annotating ion channel structures in Section 4.3 and the chapter concludes with a summary of the main findings in Section 4.4.

## 4.1 Methods

### 4.1.1 Simulations of the 5-$HT_3$ Receptor

Four structures representing three different conformational states of the 5-$HT_3$ receptor were obtained from the Protein Data Bank (PDB). Their details are listed in Table 4.1. Equilibrium MD simulations of the full length receptor were carried out following the protocol described in Section 3.3.1 and data was analysed using CHAP in version 0.9.1.[26] Further simulations were carried out on a reduced model system, comprising only the M2 helices of the 5-$HT_3$ receptor (residues D247 – T272). This reduced system was also used to probe the influence of harmonic position restraints, which in an additional set of simulations were systematically reduced compared to the prescription in Section 3.3.1. Three independent simulation repeats were carried out for each unique combination of PDB structure and simulation parameters. In the case of the complete receptor structure, simulations had a duration of 50 ns, while simulations of the M2 helix bundle, including simulations with reduced positional restraints, had a duration of 150 ns.

### 4.1.2 Simulations of the Glycine Receptor

Similar equilibrium MD simulations following the protocol from Section 3.3.1 were carried out for the glycine receptor. Table 4.2 lists the five PDB structures that were considered here, which represent three different conformational states of the channel. Again,

| PDB ID | organism | expression system | method | resolution | state |
|---|---|---|---|---|---|
| 3JAD[259] | *Danio rerio* | *Spodoptera frugiperda* | cryo-EM | 0.390 nm | closed |
| 3JAE[259] | *Danio rerio* | *Spodoptera frugiperda* | cryo-EM | 0.390 nm | open |
| 3JAF[259] | *Danio rerio* | *Spodoptera frugiperda* | cryo-EM | 0.380 nm | desensitised |
| 5CFB[257] | *Homo sapiens* | *Spodoptera frugiperda* | X-ray | 0.304 nm | closed |
| 5VDH[258] | *Homo sapiens* | *Spodoptera frugiperda* | X-ray | 0.285 nm | desensitised |

**Table 4.2:** Structures of the glycine receptor considered in this thesis. Several additional crystal structures (PDB IDs: 5VDI,[258] 5TIN,[264] and 5TIO[264]) were not considered here due to their conformational similarity with the 5VDH structure. The functional state in the final column is based on the results presented in this chapter and in each case agrees with the functional state assigned in the respective original publication.

additional simulations were performed for a structural subset of the glycine receptor, in this case comprising the M1 through M4 helices and thus the entire transmembrane (TM) domain of the channel (residues M220 – I347). As in the case of the 5-$HT_3$ receptor, simulations were conducted in triplicates for each unique combination of PDB structure and simulation parameters. Simulations of the full length receptor were run for a duration of 50 ns, while simulations of the TM domain were extended to 100 ns.

### 4.1.3 High Throughput Analysis of Ion Channel Structures

CHAP was also employed in a large-scale simulation study of hydrophobic gates in the ion channel structural proteome.[103] Equilibrium MD simulations of ∼200 different ion channel structures were carried out following the protocol described in Section 3.3.1. However, in contrast to previous simulations, the OPLS all atom force field together with united atom lipid parameters[265] was used in conjunction with the TIP4P water model[118] for this investigation. Simulations had a duration of 30 ns and three independent repeats were performed for each structure.

The structures included in this analysis form a substantial subset of the currently ∼900 ion channel structures available in the PDB (Figure 1.3). Selection was based on a resolution cutoff at 0.5 nm and low quality structures with an incomplete backbone

trace were removed from the data set. Furthermore, mutant structures were excluded from the analysis in cases where a wild-type structure of the same conformational state was available.

For each channel structure, CHAP in version 0.9.1 was used to calculate radius, hydrophobicity, and water density profiles. Barriers to water permeation were then identified from the density profile and each channel structure was thus associated with an average value of radius, hydrophobicity, and water density in the vicinity of this barrier. From these data, a phase diagram of the liquid-vapour equilibrium of nanoconfined water was constructed by locally averaging channel structures with similar pore radius and hydrophobicity over hexagonal bins.

## 4.2 Results

### 4.2.1 Hydrophobic Gating in the 5-$HT_3$ Receptor

The 5-$HT_3$ receptor is a cation-selective ion channel that is gated by the binding of the neurotransmitter 5-hydroxytryptamine (5-HT, or serotonin). In contrast to other 5-HT receptors, which belong to the class of G-protein-coupled receptors, the 5-$HT_3$ receptor is a pentameric ligand-gated ion channel (pLGIC) and a member of the family of Cys-loop receptors. Within this family, the 5-$HT_3$ receptor is most closely related to the similarly cation-selective nicotinic acetylcholine receptor.

5-$HT_3$ receptors are commonly expressed in both the central and peripheral nervous systems and are particularly prominent in the gastrointestinal tract and the brain.[266,267] They are associated with a variety of physiological functions, including roles in cognition and emesis, as well as the control of gut motility and peristalsis. Correspondingly, 5-$HT_3$ receptors are considered a potential target for drugs combatting neurological disorders such as anxiety, depression, and schizophrenia, among others.[267–269] Furthermore, 5-$HT_3$ receptor antagonists are important in the treatment of irritable bowel syndrome[270] as

well as of nausea and vomiting induced by chemotherapy.[271]

A first atomistically detailed insight into the structure of the 5-$HT_3$ receptor was provided by a crystal structure (PDB ID: 4PIR) published by Hassaine et al. [255] in 2014. While crystallised in the absence of serotonin, and thus expected to represent a closed conformation of the channel, Hassaine et al. [255] were unable to assign a functional status to this structure based on a conventional analysis of the permeation pathway using the HOLE programme. The pore radius at the putative location of the gate was found to be sufficiently large to accommodate a hydrated $Na^+$ ion. In addition, the backbone conformation of the pore lining helices more closely resembled that of the open state bacterial homologues $GLIC_O$ and GluCl than the closed state homologues $GLIC_C$ and ELIC.

It was therefore hypothesised that the 5-$HT_3$ receptor might contain a hydrophobic gate,[255] similar to the related nicotinic acetylcholine receptor, for which a hydrophobic gating mechanism had already been established.[63] Independent MD simulation studies by Trick et al. [78] and Yuan et al. [226] subsequently confirmed the existence of a hydrophobic gate in the 5-$HT_3$ receptor and established that the 4PIR structure represents a closed conformation of this channel. These results highlight the need for a new functional annotation tool that is able to go beyond an analysis of physical pore dimensions to enable the identification of hydrophobic gates.

Further studies have since yielded additional structures of the 5-$HT_3$ receptor and have elucidated different conformational states of the channel. In early 2018, Basak et al. [256] published a cryo-EM structure of the resting state (PDB ID: 6BE1), while Basak et al. [225] released cryo-EM structures of the desensitised (PDB ID: 6DG7) and open (PDB ID: 6DG8) states later in the same year. The following discussion will illustrate the utility of CHAP by directly comparing the behaviour of water confined in the pore of these four structures (Table 4.1). This will include the presentation of new data* not

*Simulations of the 5-$HT_3$ receptor presented in Trick et al. [78] are based solely on the M2 helix bundle of the 4PIR structure and were conducted by J. Trick. Results presented in this thesis are based

published in Trick et al. [78].

It also needs to be mentioned that four additional cryo-EM structures (PDB IDs: 6HIN, 6HIO, 6HIQ, and 6HIS) of the 5-$HT_3$ receptor, capturing an open and several intermediate states, were published by Polovinkin et al. [263] shortly after the release of the open and desensitised structures of Basak et al. [225]. However, due to their similarity to the data of Basak et al. [225], these structures have not been included in the present analysis. It should be noted, however, that Polovinkin et al. [263] conducted their own equilibrium MD simulations and concluded that the 9′ leucine, 13′ valine, and 17′ isoleucine residues form a hydrophobic gate, in agreement with the results presented below.

### Overall Structure of the 5-$HT_3$ Receptor

Before discussing the hydration behaviour of the 5-$HT_3$ receptor pore, it is important to become acquainted with the overall structure of the channel. As can be seen in Figure 4.1, the 5-$HT_3$ receptor follows the architectural pattern of the pLGIC family and forms a complex of five subunits in a pseudo-symmetric arrangement around a central pore, with a total molecular weight of ∼250 kDa. While there are five known subunits of the 5-$HT_3$ receptor, denoted by alphabetic suffixes as 5-HT3A through 5-HT3E, all available atomic resolution structures to date capture a homomeric assembly of five 5-HT3A subunits.[225,255,256,263]

Each individual subunit comprises three domains, an extracellular domain (ECD), where the ligand-binding site is located, an $\alpha$-helical transmembrane domain (TMD), and an intracellular domain (ICD), which is not fully resolved in the available structures. The TMD comprises four $\alpha$-helices, M1 through M4, where the M2 helices face the transmembrane pore, while the other helices separate the bundle of M2 helices from

on an independent set simulations of the full length 5-$HT_3$ receptor conducted by G. Klesse. Since the work of Trick et al. [78] predates the development of CHAP, there are also some methodological differences in the details of the simulation protocol. However, this does not affect conclusions drawn about the functional state of the 4PIR structure.

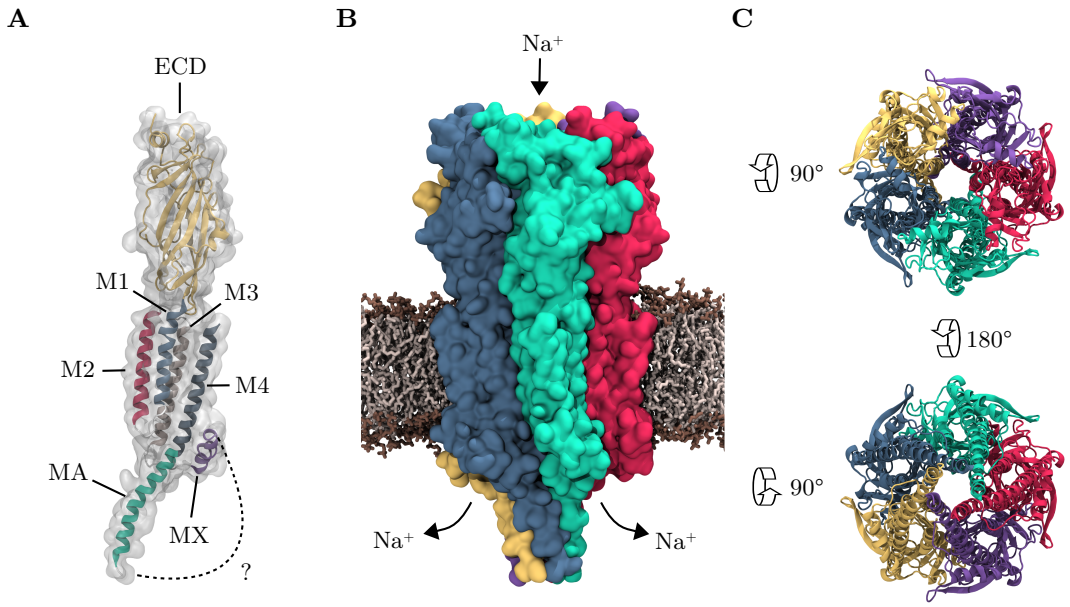



**Figure 4.1:** Structural architecture of the 5-$HT_3$ receptor. (A) Cartoon representation of a single subunit, where the different transmembrane and intracellular helices as well as the extracellular domain (ECD) are highlighted in separate colours. The segment connecting the MX and MA helices missing from the structure is represented by a dashed line. (B) Surface representation of the complete 5-$HT_3$ receptor structure coloured by subunit. A DOPC bilayer membrane is shown in licorice representation. $Na^+$ ions are believed to enter the central pore of the channel on the extracellular side, but leave through lateral portals after passing the transmembrane domain. (C) Five-fold pseudosymmetric arrangement of subunits around a central pore as seen from the extracellular (top) and intracellular (bottom) domains. The structure shown represents the closed conformation (PDB ID: 4PIR) of the channel.

the membrane. In the intracellular domain, the central pathway is lined by the MA helices. However, it is believed that ions entering the central pore from the extracellular space leave the channel through lateral portals located below the MX helices on the intracellular side of the membrane. Thus, ions may never actually pass through the intracellular part of the central pore and the channel gate is located on the M2 helix.

### A Hydrophobic Constriction in the Transmembrane Pore

Figure 4.2 compares the dimensions and hydrophobicity of the central pore through the 5-$HT_3$ receptor across four experimentally determined structures (Table 4.1). The corresponding scalar radius and hydrophobicity profiles are shown in Figure 4.3.

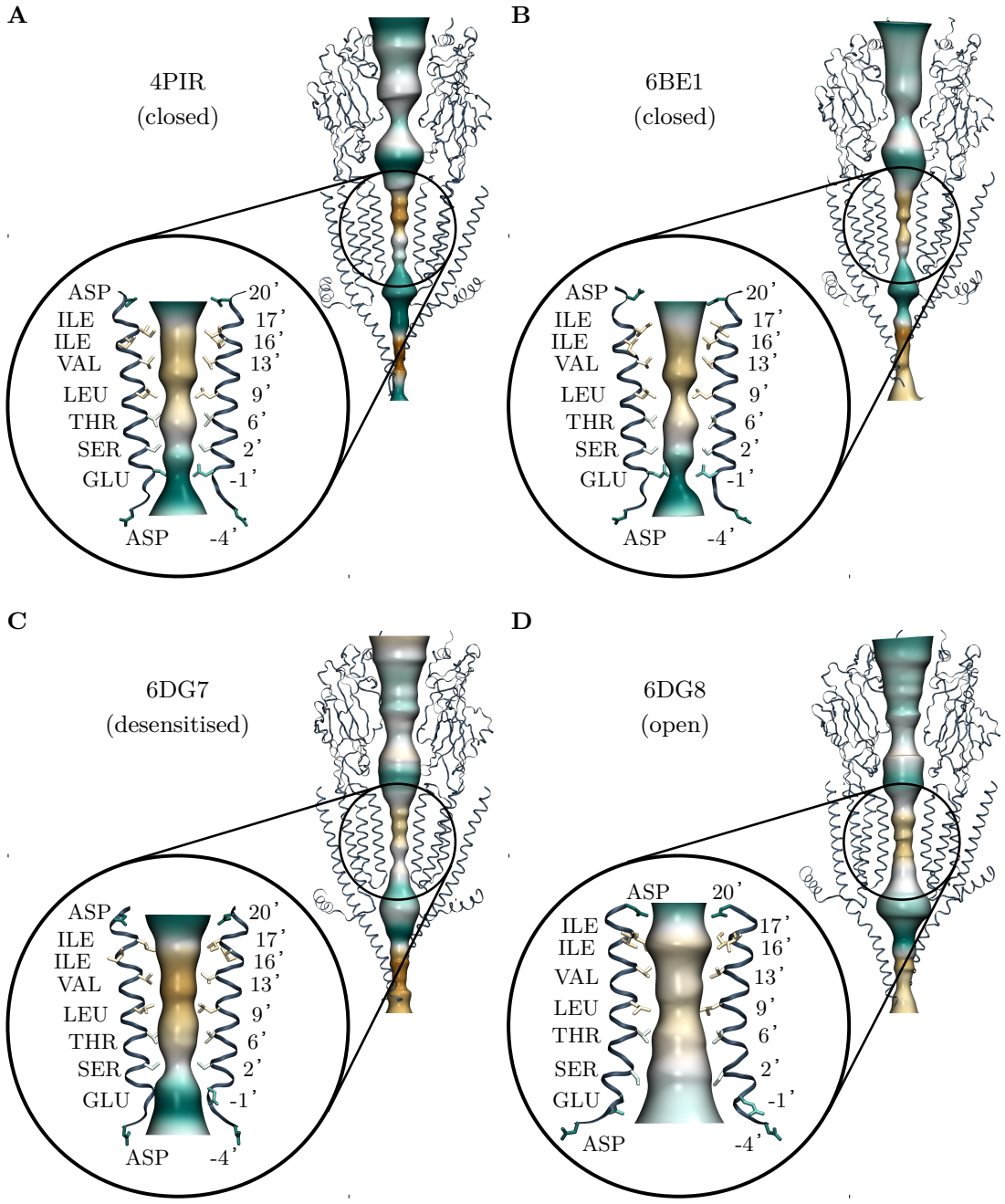



**Figure 4.2:** Size and hydrophobicity of the central pore in the 5-$HT_3$ receptor. (A) – (D) Hydrophobicity coloured pore through four different structural conformations of the channel. Two of the receptor subunits are shown as ribbons. A magnified extract shows details of the transmembrane pore lined by the M2 helices, with pore-facing sidechains shown in licorice representation and coloured by hydrophobicity.

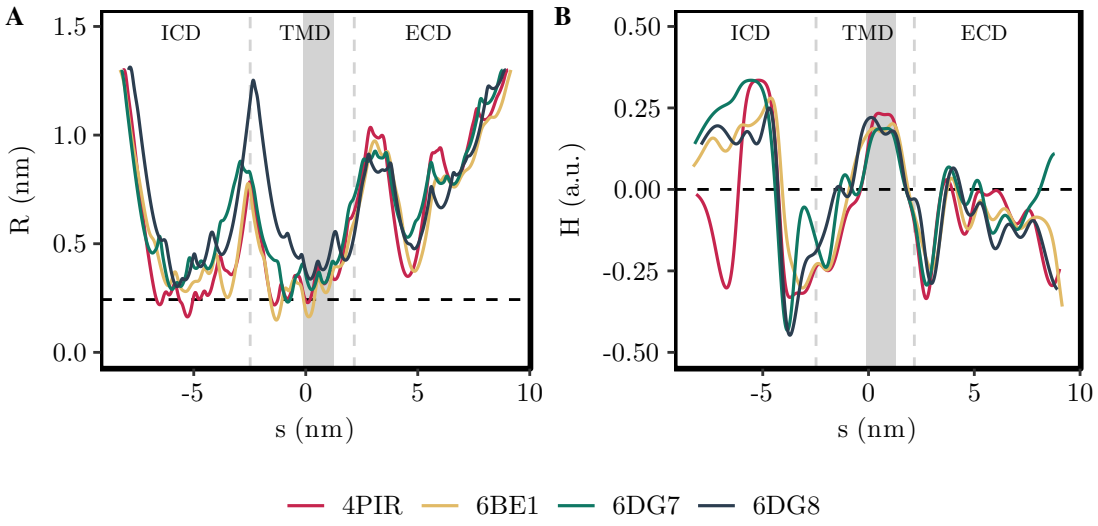



**Figure 4.3:** Radius and hydrophobicity profiles for four structures of the 5-$HT_3$ receptor. (A) Pore radius profile. The dashed black line at 0.243 nm indicates the radius of a hydrated $Na^+$ ion.[272] (B) Hydrophobicity profile based on the amino acid hydropathy scale by Wimley and White [238]. Vertical dashed lines mark the limits of the intracellular (ICD), transmembrane (TMD), and extracellular (ECD) domains, and the shaded background indicates the location of the hydrophobic gate.

It is immediately apparent that in the absence of the open and desensitised state structures (PDB IDs: 6DG7 and 6DG8) as a reference point, the original closed state structure (PDB ID: 4PIR) presents a challenge for conventional size-based approaches to functional annotation (Figure 4.2A). The radius of the pore does not vary substantially throughout the transmembrane segment and even the constrictions near the $-1'$ glutamate and $9'$ leucine residues are only slightly narrower than a hydrated $Na^+$ ion (Figure 4.3A). There is thus no obvious steric impediment for ion permeation in this structure.

However, while the intracellular half of the transmembrane pore is lined by hydrophilic and charged residues, there is a highly hydrophobic segment in its extracellular half, spanning from the $9'$ leucine to $16'$ isoleucine residues (Figure 4.3B). In this region, the pore radius ranges from ∼0.23 nm near the $9'$ leucine residues to ∼0.30 nm in the vicinity of the $16'$ isoleucines. This section of the pore is thus smaller than the critical radius

below which liquid-vapour oscillations occur in model nanopores (Figure 1.4). As will be shown below, the pore does indeed exhibit dewetting in this region, which thus forms a hydrophobic gate that renders the channel nonconducting.

While its overall fold is similar, the cryo-EM structure of the closed state (PDB ID: 6BE1) exhibits slightly more narrow constrictions near the $-1'$ glutamate and $9'$ leucine residues (Figure 4.3A). However, this structure was determined at a resolution of only 0.431 nm, while the X-ray structure (PDB ID: 4PIR) achieved a resolution of 0.350 nm and its radius profile is thus likely more reliable.

By comparison, the transmembrane pore widens in the other two structures (PDB IDs: 6DG7 and 6DG8). In the desensitised state (PDB ID: 6DG7), the narrowest constriction is formed by the $2'$ serine residues and at a radius of ∼0.24 nm is large enough to permit the passage of a hydrated $Na^+$ ion (Figure 4.3A). In addition, the hydrophobic constriction near the $9'$ leucine residues is dilated compared to the closed state, but its radius of ∼0.30 nm still falls into the range where dewetting may occur. In the open conformation, the pore is dilated further still, with its narrowest constriction of ∼0.32 nm formed by the $9'$ leucine residues.

It is also evident from Figure 4.2 that there is an additional hydrophobic constriction in the intracellular part of the central pore. While very narrow in the X-ray structure of the closed state (PDB ID: 4PIR), this segment of the pore is sufficiently large to accommodate a hydrated $Na^+$ ion in the other three structures (Figure 4.3A). However, as ions are believed to bypass this constriction by laterally exiting the channel below the transmembrane domain (Figure 4.1), the intracellular domain does not act as a gate and is therefore not the focus of this study.

### Dewetting of the Hydrophobic Gate

Figure 4.4 depicts the behaviour of water in MD simulations of the 5-$HT_3$ receptor. As can be seen in Figure 4.4A, an approximately 1 nm long stretch of the transmembrane

pore, corresponding to the hydrophobic segment between 9′ leucine and 16′ isoleucine, is devoid of water in both closed state structures (PDB IDs: 4PIR and 6BE1), as well as in the desensitised conformation (PDB ID: 6DG7). The average water density in the hydrophobic gate region is near zero in these three structures, consistently across three simulation repeats (Figure 4.4B).

This indicates that these structures correspond to nonconducting functional states of the channel, in which a hydrophobic gate prevents the flow of ions. Importantly, this functional annotation was not evident a priori on the basis of structural data alone, as neither the 4PIR nor the 6DG7 structure possess a sterically occluded transmembrane pore. However, it needs to be pointed out that while the behaviour of water can be used to discriminate between conducting and nonconducting states, it cannot be used to establish which state of a gating mechanism a structure corresponds to. The labels "closed" and "desensitised" are assigned on the basis of the respective experimental conditions (the 6DG7 structure was determined in the presence of serotonin), as well as on a comparison of the radius profile and pore conformation with other members of the pLGIC family.[225]

By contrast, in the open state structure (PDB ID: 6DG8), the transmembrane pore remains fully hydrated so that an aqueous pathway connects the extracellular and intracellular domains (Figure 4.4A). Correspondingly, the average water density in the transmembrane domain of this conformation of the receptor remains close to the density of bulk water in three independent simulations (Figure 4.4B). These results also highlight how the switch-like nature of the liquid-vapour equilibrium of nanoconfined water (Figure 1.4) is exploited in hydrophobic gating: a relatively minor increase of pore radius in the open state structure (PDB ID: 6DG8) compared to the desensitised conformation (PDB ID: 6DG7) is sufficient to convert an entirely dewetted pore into a fully hydrated pore.

It is also worthwhile to consider the behaviour of water in the intracellular domain.

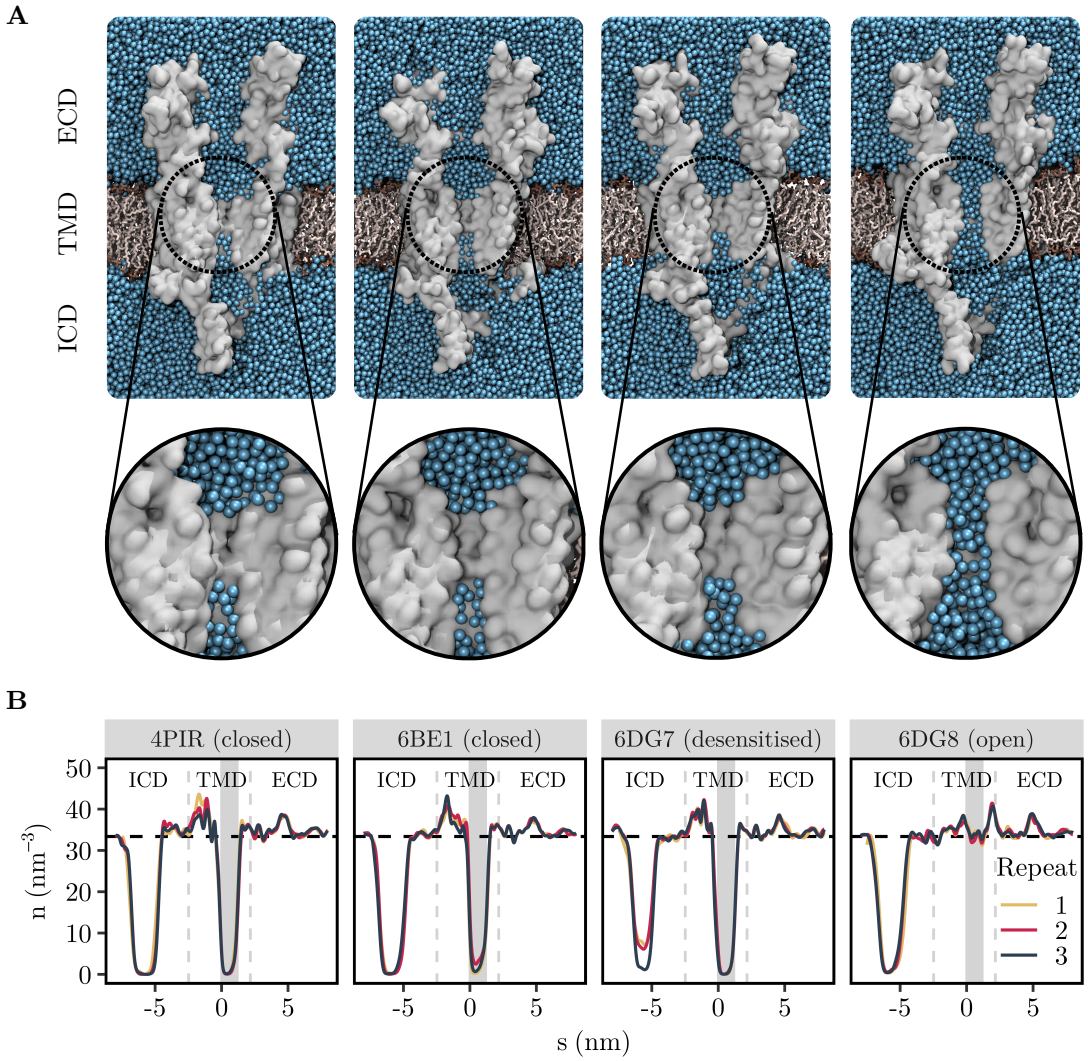



**Figure 4.4:** Hydration of the 5-HT$_3$ receptor pore in equilibrium MD simulations. In the closed and desensitised states, the hydrophobic region in the transmembrane pore of the channel remains dewetted, while in the open state the extracellular and intracellular domains are connected by an aqueous pathway. (A) Snapshots from the end of a 50 ns long MD simulation of four conformational states of the 5-HT$_3$ receptor. The channel protein is shown in surface representation, where the two reader-facing subunits have been omitted to permit an unobscured view of the permeation pathway. Membrane lipids are shown in licorice representation and water oxygens are represented by their van der Waals spheres. (B) Time-averaged water density profiles derived from three independent simulations of each conformational state of the receptor. Vertical dashed lines denote the boundary between the intracellular (ICD), transmembrane (TMD), and extracellular (ECD) domains and the shaded background indicates the location of the hydrophobic gate between the 9′ leucine and 17′ isoleucine resides. The dashed horizontal line indicates the density of bulk water of 33.37 nm$^{-3}$.

As shown in Figure 4.4, this part of the central pore dewets in all four structures of the 5-$HT_3$ receptor, including the open conformation (PDB ID: 6DG8). At the same time, an intracellular vestibule just below the transmembrane pore is connected to the cytosolic bulk water domain through an aqueous lateral pathway (Figure 4.4A). This observation strengthens the view that permeating ions take this sideways route and thus avoid the hydrophobic constriction in the intracellular central pore.

### Influence of the Structural Model

While recent advances in high performance computing have greatly extended the time and length scales over which MD simulations are feasible, the relatively large size of ion channel proteins still poses a challenge. For example, the 5-$HT_3$ receptor considered here is ∼16.5 nm long and has an external diameter of ∼8 nm. Consequently, a typical simulation system consists of more than 300 000 particles (Figure 4.5A). For even larger ion channels, such as ryanodine receptors[273] or the mechanosensitive channel Piezo,[31] atomistically resolved simulation systems may approach 1 000 000 particles, leading to increasingly long simulation times[†].

This is particularly vexing, as in many cases only a small region of the channel protein is of interest. For example, in many ion channels the gate is located in the transmembrane domain and frequently it is even possible to pinpoint the exact residues involved in channel gating, e.g. through electrophysiological mutagenesis experiments. In an effort to limit computational resource use, some prior simulation studies of hydrophobic gating have therefore examined the ion permeation pathway by focussing on a minimal subset of the ion channel structure, typically comprising only the pore-lining $\alpha$-helices,[63,78] while disregarding the extracellular and intracellular domains entirely.

However, this procedure raises the concern that a reduced structural model may im-

[†]In theory, the computational cost of an MD simulation employing SPME electrostatics increases superlinearly with the number of particles, $N$, according to $\mathcal{O}(N \log N)$. However, in a practical implementation, the finite size of main and GPU memory, as well as limited interconnect throughput rates may impose even stricter resource constraints.

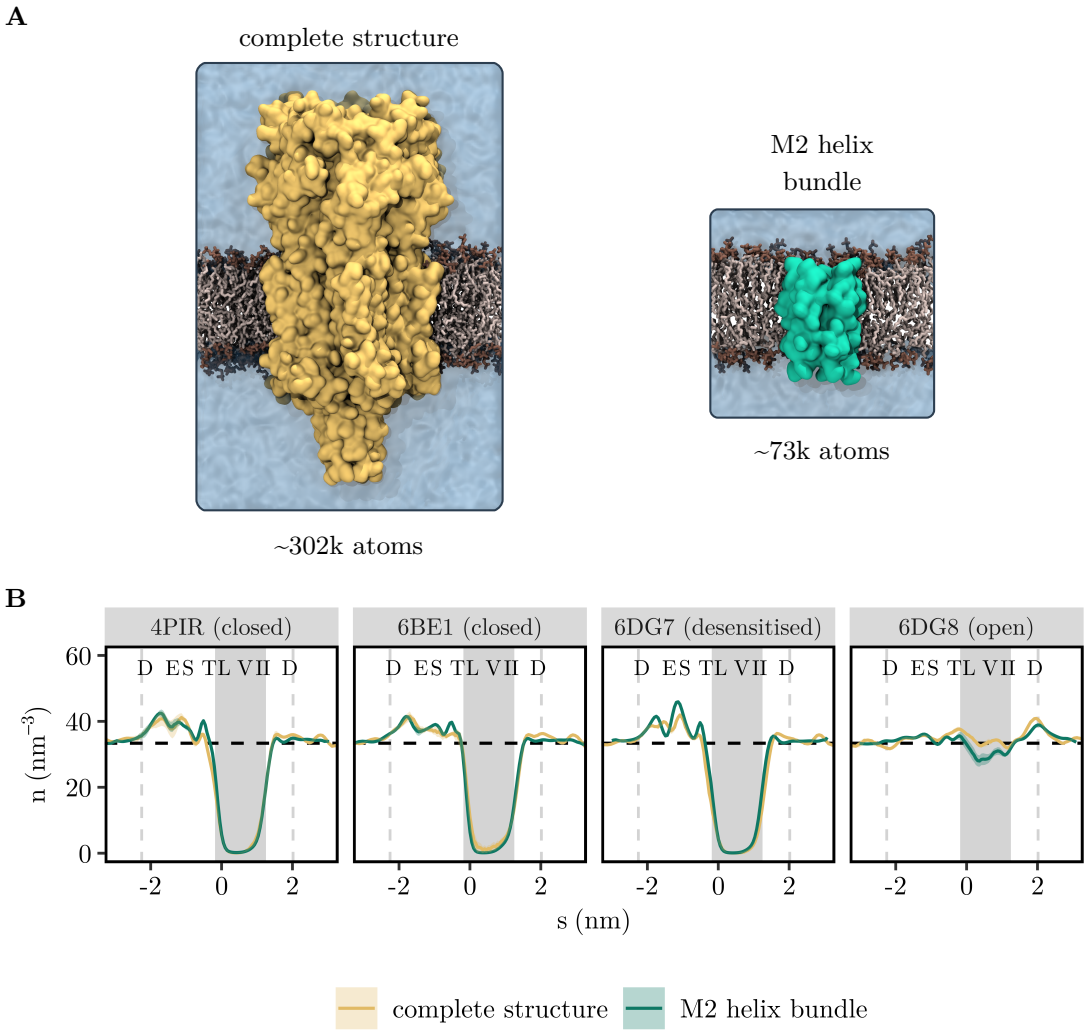



**Figure 4.5:** Influence of 5-HT$_3$ receptor structural model. (A) Simulation system based on the complete 5-HT$_3$ receptor structure compared to a much smaller simulation system based on the only the M2 helix bundle. The latter system comprises a factor of ∼4 fewer particles. Structures shown represent the closed (PDB ID: 4PIR) conformation of the channel. (B) Comparison of time-averaged water density profiles in the transmembrane domain of four different conformations of the 5-HT$_3$ receptor. Vertical dashed lines indicate the extent of the M2 helix bundle pore and the shaded background represents the hydrophobic gate region, while the dashed horizontal line represents the density of bulk water of 33.37 nm$^{-3}$. Confidence bands represent the standard deviation over three independent MD simulations.

pact the hydration equilibrium inside the channel pore, i.e. it may artificially bias the channel towards either the wetted or dewetted state, thus leading to an inaccurate func-

tional annotation. For example, omitting parts of the channel structure containing a significant number of charged residues may alter the electrostatic environment in the interior of the channel. Moreover, removing outlying transmembrane helices may open up nonphysiological fenestrations that permit lipid tails to access the pore. This could in turn contribute to the expulsion of water molecules from the channel, either by increasing the local hydrophobicity or directly through steric clashes.

To test whether a reduced model can accurately capture the hydration state of a channel pore, an additional set of simulations of the 5-$HT_3$ receptor, including all four previously considered structural conformations, was conducted. In these simulations, only the M2 helices of the channel, embedded in a lipid bilayer and solvated in a NaCl electrolyte, were simulated (Figure 4.5A). Compared to the full length 5-$HT_3$ receptor, this reduced simulation system contains ∼4 times fewer particles and permits significant savings in simulation time. While in the closed and desensitised conformations the M2 helices form a tightly packed bundle, some lateral openings exist in the open conformation. Interference from the bilayer was therefore prevented through additional restraints on the lipid molecules (see Section 4.1 for details).

As can be seen in Figure 4.5B, the resulting water density throughout the transmembrane domain is nearly identical between the two different simulation systems. In both the closed (PDB IDs: 4PIR and 6BE1) and desensitised (PDB ID: 6DG7) conformations that have previously exhibited significant dewetting (Figure 4.4), water still favours the vapour state when only the M2 helices are simulated. In the open conformation (PDB ID: 6DG8), the time-averaged water density in the hydrophobic gate region of the M2 helix bundle falls slightly below the value observed for the full length structure, reflecting the fact that the M2 helix bundle intermittently dewets. Nevertheless, the average water density remains close to the bulk value, which suggests that the influence of the structural model on the hydration equilibrium is likely limited and is not sufficiently strong to bias an otherwise hydrated channel towards the vapour state.

### Influence of Positional Restraints

An important aspect of the simulation protocol introduced in Section 3.3.1 is that the conformation of the simulated channel protein is maintained close to the original experimental structure by means of harmonic position restraints. Without these restraints, a channel might begin to undergo allosteric transitions between different functional states, rendering any functional annotation derived from such a simulation more complicated to interpret. Open state channel structures in particular could even collapse into an nonphysiological conformation with a sterically occluded pore.

However, in addition to controlling the conformation of the channel protein, the restraints will also influence the local flexibility of the pore surface and might thus influence the behaviour of water molecules confined inside the pore. From earlier simulation studies of both model pores[44] and carbon nanotubes,[274] it is known that increased surface stiffness can counterbalance the effect of pore surface hydrophobicity and bias the hydration equilibrium towards the liquid phase. In agreement with this, combined theoretical and simulation studies have more recently reported a strong dependence of the rate of evaporation of water confined between hydrophobic plates on the flexibility of the confining material.[275,276]

The impact of harmonic restraints on the behaviour of water confined in ion channel pores was therefore investigated through additional simulations, in which the magnitude of the restraining force was systematically varied between very soft restraints of $10\,\mathrm{kJ\,mol^{-1}\,nm^{-2}}$ and the strong restraints of $1000\,\mathrm{kJ\,mol^{-1}\,nm^{-2}}$ applied in all previous simulations. In addition, a protocol where restraints are applied only to C$\alpha$ atoms is compared to a protocol in which all heavy atoms, including those in amino acid side chains, are restrained. Following the results presented in the previous section, the pore was represented by the M2 helix bundle of the closed conformation of the 5-$HT_3$ receptor (PDB ID: 4PIR).

Figure 4.6A shows how the time-averaged RMSD of protein backbone and side chain

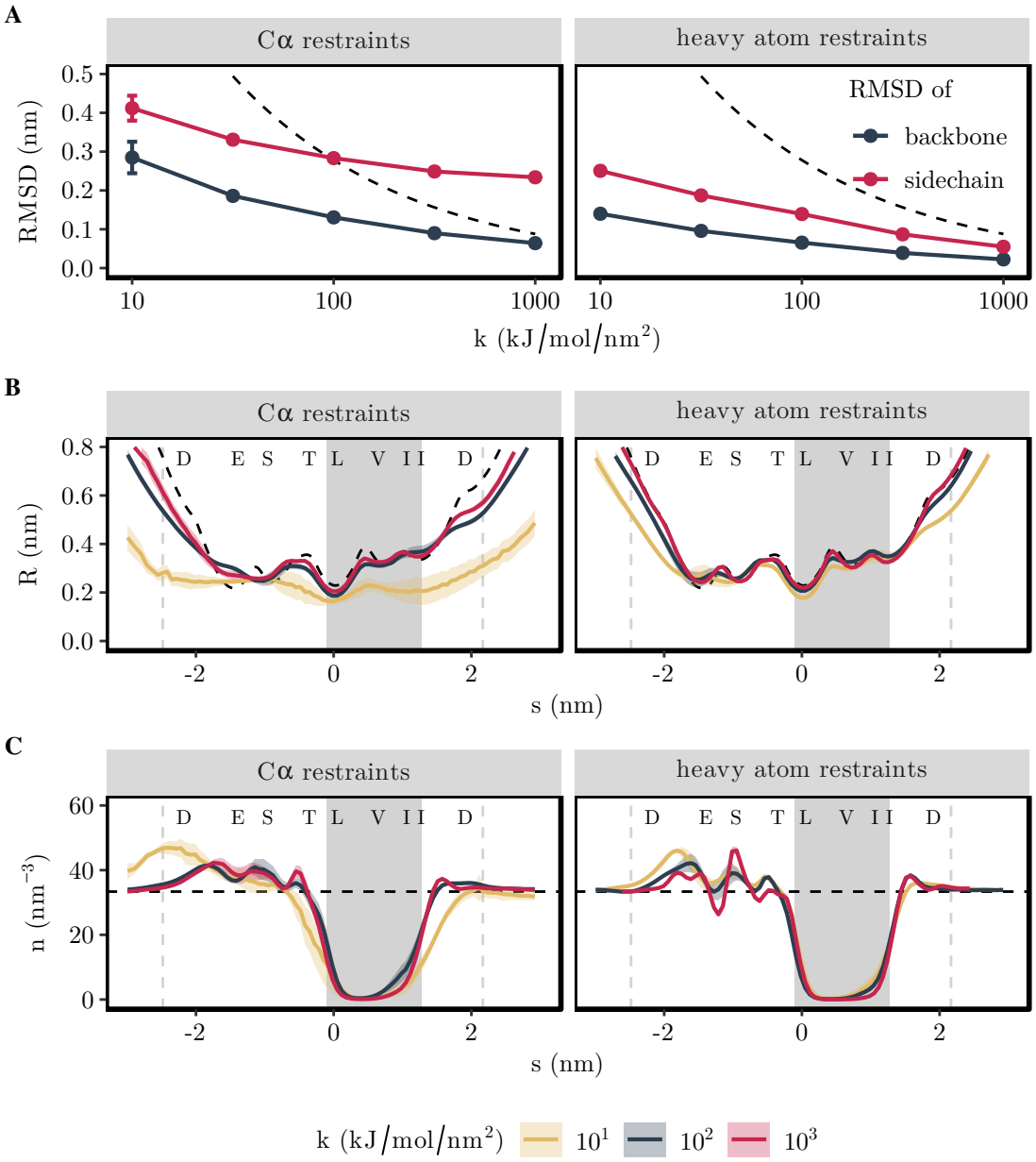



**Figure 4.6:** Influence of positional restraints in simulations of the M2 helix bundle of the closed state (PDB ID: 4PIR) 5-$HT_3$ receptor. (A) The time-averaged RMSD of both backbone and sidechain atoms decreases as the restraining force is increased. Error bars denote the standard deviation across three independent repeats. The dashed line indicates the RMSD expected for an individual harmonic oscillator. (B) Pore radius profile. The dashed line indicates the radius profile calculated from the crystal structure. (C) Water density profile. The dashed horizontal line indicates the density of bulk water of 33.37 $nm^{-3}$. In panels (B) and (C), shaded background and vertical dashed lines indicate the position of the hydrophobic gate and the extent of the protein respectively. Single letter codes for the pore-facing amino acids are given at the top of each panel.

atoms respectively varies with the strength of the restraining force. In the case of C$\alpha$ restraints, the RMSD decreases from $\sim$0.3 nm to $\sim$0.1 nm for backbone atoms and from $\sim$0.4 nm to $\sim$0.3 nm for side chain atoms, as the restraining force increases. By comparison, the crystal structure of the 5-$HT_3$ receptor has a resolution of $\sim$0.35 nm, and the resolution of the cryo-EM structures of this channel ranges from $\sim$0.33 nm to $\sim$0.43 nm (Table 4.1). Therefore, a restraining force of $\geq$100 kJ mol$^{-1}$ nm$^{-2}$ can be considered sufficient to ensure that a simulated channel accurately represents an experimentally determined structure.

As can be seen from Figure 4.6B, the radius profile of the simulated channel generally agrees well with that of the original structure, provided that the restraining force exceeds $\sim$100 kJ mol$^{-1}$ nm$^{-2}$. Under this condition, deviations from the experimental radius profile do not typically exceed $\sim$0.05 nm, except near the openings of the M2 helix bundle, where side chains can rotate more freely. By contrast, under weak restraints the radius profile no longer resembles the experimental profile and the pore becomes significantly narrower. The large magnitude of the deviations shown in Figure 4.6B is due to the fact that structural elements of the channel that anchor the pore-lining helices in their physiological position were absent from these simulations, which comprised only the M2 helix bundle itself.

Importantly, as shown in Figure 4.6C, dewetting of the hydrophobic gate region can be observed irrespective of the strength of the restraining force. While the water density profile for weak restraints ($k = 10$ kJ mol$^{-1}$ nm$^{-2}$) deviates somewhat from the profiles obtained with stronger restraints, the deviations are most pronounced outside the hydrophobic gate region and are most likely due to the differences in radius profile noted above. Consequently, the application of harmonic restraints does not appear to have a strong influence on the hydration behaviour of the pore for the typical dimensions of an ion channel. It cannot be ruled out that an effect would be observed if the restraints were increased even further. However, a restraining force significantly larger

than $\sim$1000 kJ mol$^{-1}$ nm$^{-2}$ would lead to dynamic instabilities and is therefore not commonly applied in MD simulations.

These conclusions do not change when restraints are applied to all heavy atoms rather than only to C$\alpha$ atoms. Unsurprisingly, the RMSD of side chain atoms decreases significantly in this case and falls below $\sim$0.1 nm for the strongest restraints (Figure 4.6A). Correspondingly, the radius profile of the simulated protein resembles that of the experimental structure very closely, even for comparably weak restraints (Figure 4.6B). As shown in Figure 4.6C, the hydrophobic gate region dewets irrespective of the strength of the restraining force, as it had when C$\alpha$ restraints were used.

The choice of C$\alpha$ restraints over heavy atom restraints in the simulation protocol suggested in Section 3.3.1 is thus motivated primarily by the fact that side chain rotamer conformations in experimentally determined protein structures of modest resolution are frequently less certain than the positions of backbone atoms. Moreover, structural models are often built under an assumption of subunit pseudo-symmetry, i.e. $C_n$-symmetry, while under physiological conditions side chains are likely able to reorient themselves independently of one another. It therefore appears more sensible to preserve the orientational degrees of freedom of side chains in an MD simulation rather than to restrain side chains in a potentially arbitrary conformation.

### 4.2.2 Hydrophobic Gating in the Glycine Receptor

Glycine receptors are another member of the pLGIC family, but in contrast to 5-HT$_3$ receptors they are selective for $Cl^-$ ions and act as a mediator for inhibitory rather than excitatory neurotransmission.[277] Most prominently expressed in the spinal chord and brain stem, glycine receptors are involved in the physiology of motor coordination, sensory processing, and pain perception,[278,279] and have received attention as a therapeutic target.[280] Moreover, mutations of the glycine receptor are responsible for hyperekplexia, a rare genetic disorder also known as startle disease.[281]

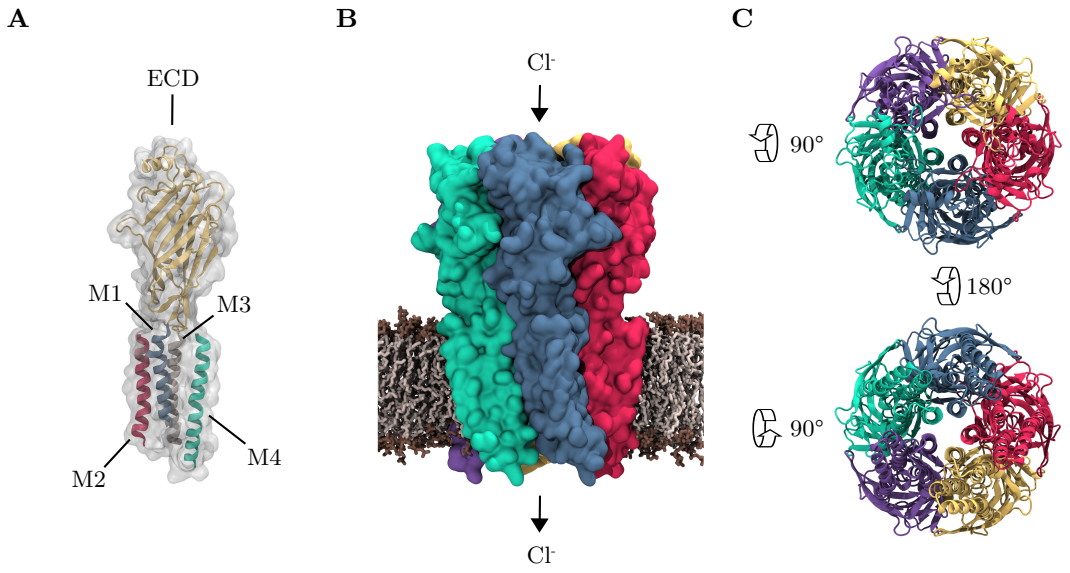



**Figure 4.7:** Structural architecture of the glycine receptor. (A) Cartoon representation of a single subunit, where the different transmembrane helices and the extracellular domain (ECD) are highlighted in separate colours. (B) Surface representation of the complete glycine receptor structure coloured by subunit. A DOPC bilayer membrane is shown in licorice representation. The glycine receptor is $Cl^-$ selective and ions are conducted through a central pore enclosed by the five subunits. (C) Five-fold pseudosymmetric arrangement of subunits around a central pore as seen from the extracellular (top) and intracellular (bottom) domains. The structure shown represents the closed conformation (PDB ID: 3JAD) of the channel.

### Overall Structure of the Glycine Receptor

Atomistically resolved structures of the glycine receptor first became available in 2015, when Du et al. [259] released three cryo-EM structures of the ion channel in its closed, open, and desensitised conformations. These data were supplemented by an X-ray structure of the closed state solved by Huang et al. [257] shortly thereafter, and by additional X-ray structures of the desensitised state in 2017.[258,264] Notably, the functional status of these structures was not ambiguous as in the case of the first 5-$HT_3$ receptor structure. This allowed Trick et al. [78] to conclude that using the behaviour of water as a proxy for functional annotation would yield reasonable results when comparing different conformational states of the same channel.

However, it needs to be pointed out that the cryo-EM structure of the open state

presented by Du et al. [259] is the subject of an ongoing debate about its physiological relevance. Several authors have noted that the pore radius of this structure exceeds the pore radius of the open state structures of other anion-selective pLGICs.[260–262] Furthermore, the collapse of this structure in MD simulations[261,262] as well as its permeability by polyatomic anions[261] suggest that it does not capture the true open state of the glycine receptor. This issue will be discussed further in Section 4.3.

Figure 4.7 illustrates the overall architecture of the glycine receptor. Five subunits form a pseudosymmetric arrangement around a central pore, which forms the ion permeation pathway. As in the 5-$HT_3$ receptor (Figure 4.1), four $\alpha$-helices (M1 – M4) form the transmembrane domain, but the construct used for structure determination does not include an intracellular domain and the MA and MX helices are absent from this structure. However, an extracellular domain is connected to the M1 helix and contains the ligand-binding site.

The following sections will illustrate the utility of CHAP in functionally annotating the available glycine receptor structures. To this end, equilibrium MD simulations of water were carried out[‡] for five different structures of the channel in three different functional states, derived from both cryo-EM and X-ray crystallography experiments (Table 4.2). It will thus be demonstrated that this approach is generally applicable and yields results in agreement with the known pharmacological properties of ligands used to stabilise the various conformations of the receptor.

### Hydrophobic Constrictions in the Transmembrane Pore

Figure 4.8 shows a representation of the ion conducting pore in the five structures of the glycine receptor listed in Table 4.2. A magnification highlights the M2 helices, which

[‡]Simulations of the cryo-EM structures of the glycine receptor (PDB IDs: 3JAD, 3JAE, and 3JAF) presented in Trick et al. [78] were conducted by S. Chelvaniththilan, while simulations of the X-ray structure (PDB ID: 5CFB) were conducted by G. Klesse. Since the publication of that paper, the development of CHAP has yielded several methodological refinements and the results presented in this thesis are thus based on an independent set of simulations conducted by G. Klesse, which also include an additional X-ray structure (PDB ID: 5VDH).

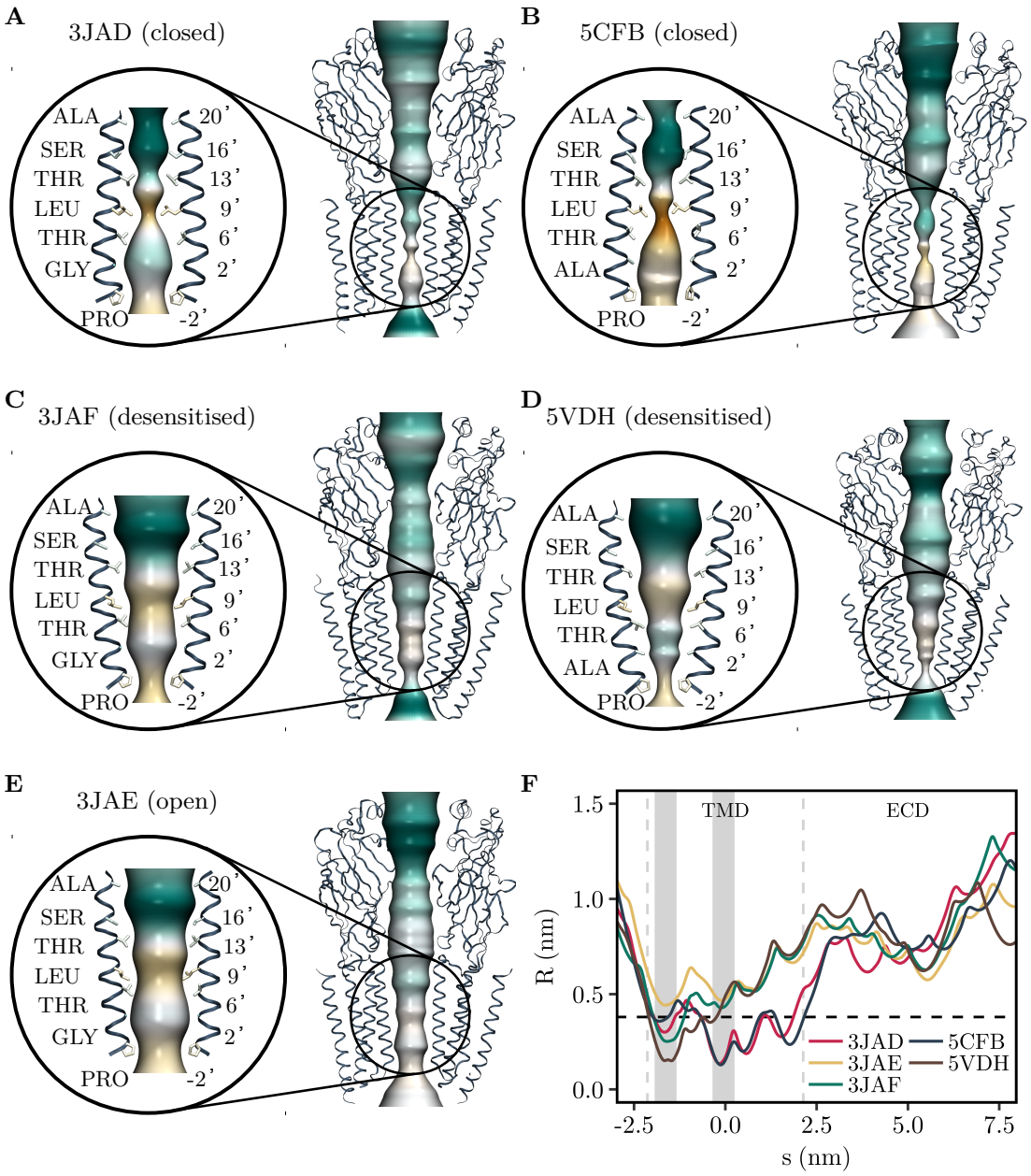



**Figure 4.8:** Size and hydrophobicity of the ion conduction pathway through the glycine receptor. (A) – (E) Hydrophobicity coloured pore through five different structural conformations of the channel. Two of the receptor subunits are shown as ribbons. A magnified extract shows details of the transmembrane pore lined by the M2 helices, with pore-facing sidechains shown in licorice representation and coloured by hydrophobicity. (F) Pore radius profiles for all five structures. Vertical dashed lines indicate the extent of the transmembrane domain (TMD), while the dashed horizontal line represents the radius of a hydrated $Cl^-$ ion of 0.38 nm. The shaded background indicates the location of the leucine and proline gates, which form the narrowest constriction in the closed and desensitised conformations respectively.

line the transmembrane segment of the pore. Note that in these helices the amino acid sequences of glycine receptors from *Danio rerio* (PDB IDs: 3JAD, 3JAE, and 3JAF) and *Homo sapiens* (PDB IDs: 5CFB and 5VDH) are very similar and differ only in the $2'$ position, where the human ion channel contains an alanine in contrast to the glycine present in the zebrafish receptor.

As can be seen from Figure 4.8, in the extracellular domain the pore is always wide enough to allow ion permeation, but in the closed and desensitised states of the channel the arrangement of the M2 helices can create narrow, hydrophobic constrictions near the $9'$ leucine and $-2'$ proline residues respectively. Both residues are highly conserved and mutations at these sites have been linked to hyperekplexia.[281]

In both the cryo-EM (PDB ID: 3JAD) and X-ray (PDB ID: 5CFB) structures of the closed state, the $9'$ leucine ring constricts the pore to a radius of $\sim$0.14 nm, thus sterically obstructing the passage of $Cl^-$ ions, whose dehydrated radius is 0.18 nm. The pore radius at the constriction site widens significantly in the desensitised conformation so that even a fully hydrated $Cl^-$ ion with a radius of $\sim$0.38 nm can pass it without steric clashes. However, in the desensitised state, the $-2'$ proline residues form a constriction with a radius of $\sim$0.14 nm in the X-ray structure (PDB ID: 5VDH) and $\sim$0.25 nm in the cryo-EM structure (PDB ID: 3JAF). Notably, both constriction sites are dilated in the open state structure (PDB ID: 3JAE).

### Dewetting of the Leucine and Proline Gates

The hydration behaviour of the glycine receptor pore in equilibrium MD simulations is illustrated in Figure 4.9. As can be seen, in all five structural conformations, nearly the entire permeation pathway is consistently filled with water (Figure 4.9A–E) and water density remains near the bulk value throughout most of the pore (Figure 4.9F). Only short stretches of the transmembrane pore exhibit dewetting in simulations of the closed and desensitised state structures.

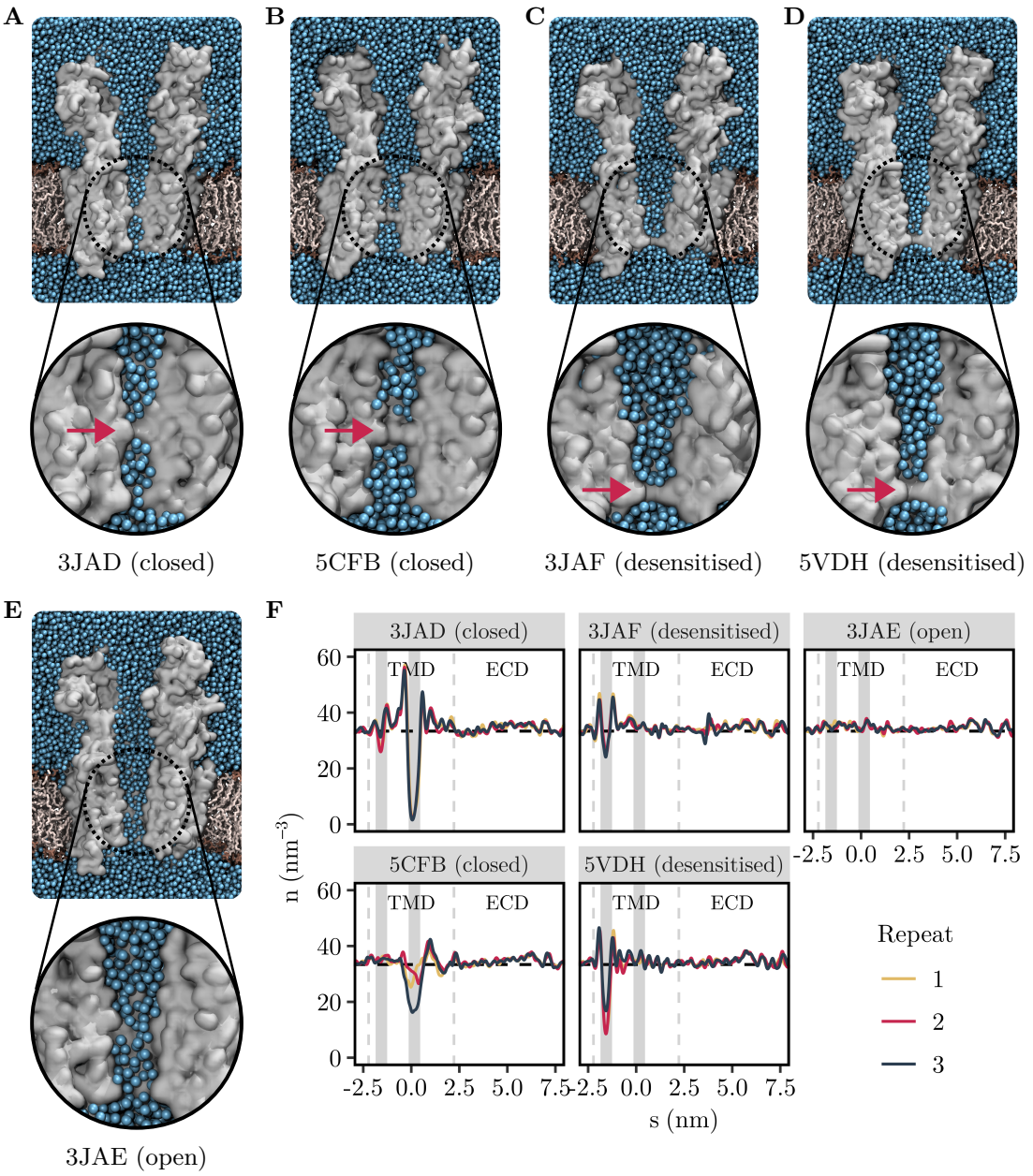



**Figure 4.9:** Hydration of the glycine receptor pore in equilibrium MD simulations. (A) – (E) Snapshots from the end of a 100 ns long simulation for five different structural conformations of the channel. The receptor is shown in surface representation, where the two reader-facing subunits have been omitted. Bilayer lipids are shown in licorice representation and water oxygens are represented by blue van der Waals spheres. Red arrows indicate the location of the dewetted leucine and proline gates in the closed and desensitised conformations of the channel respectively. (F) Water density profiles calculated with CHAP. Background shading indicates the position of the $-2'$ proline and $9'$ leucine gates, while the vertical dashed lines mark the extent of the transmembrane domain.

In both the cryo-EM (PDB ID: 3JAD) and X-ray (PDB ID: 5CFB) structures of the closed state, dewetting can be observed near the $9'$ ring of leucine residues (Figure 4.9A–B). This corresponds to a decrease in average water density to near zero in the case of the cryo-EM structure, while in the X-ray structure, water may intermittently enter the pore, resulting in an average water density below that of bulk water, but still significantly above zero (Figure 4.9F). For the 5CFB structure, there is also a greater spread across repeats, although all three repeats exhibit some degree of dewetting near the $9'$ leucine residues. Thus, equilibrium MD simulations of water indicate that these two structures correspond to a nonconducting functional state of the channel and suggest that the gate is formed by the $9'$ residues, in agreement with functional data.[282]

By contrast, the pore of the desensitised state structures (PDB IDs: 3JAF and 5VDH) remains fully hydrated near the $9'$ site (Figure 4.9C–D). However, water is partially expelled from the constriction formed by the $-2'$ proline residues, even though the average water density does not drop to zero in this region (Figure 4.9F). Remarkably, both desensitised structures also exhibit two peaks in water density just outside the proline gate. These are due to the fact that the permeation pathway is so narrow in this region, that the distance between water molecules and the pore surface decreases to below the typical intermolecular distance of liquid phase water. Nevertheless, equilibrium simulations not only confirm that the $-2'$ proline residues form a desensitisation gate, but also demonstrate that the $9'$ leucine residues do not constitute an obstacle to permeation in this conformation of the channel.

Figure 4.9E shows that neither the $9'$ leucine nor the $-2'$ proline site exhibit dewetting in simulations of the open state glycine receptor structure (PDB ID: 3JAE). The large radius of the pore in this conformation exceeds the critical radius for evaporation and consequently liquid water with a bulk-like density is present throughout the permeation pathway (Figure 4.9F). It should also be noted that the hydrophobic segment in the glycine receptor pore is shorter than in the 5-$HT_3$ receptor pore, which according to

Equation 1.4 implies that this critical radius itself will be smaller as well.

### Influence of the Structural Model

In Section 4.2.1, it was shown that in simulations of the 5-$HT_3$ receptor, water confined in the channel pore behaved similarly regardless of whether the full channel or a structural subset comprising its M2 helix bundle was simulated. Simulations of a reduced structural model potentially provide an avenue towards faster and less computationally expensive functional annotation. It is therefore interesting to check whether the pore hydration equilibrium is similarly independent of distal parts of the channel structure in the case of the glycine receptor as well.

Unfortunately, preliminary simulations indicated that a structural model comprising only the M2 helix bundle are not sensible in the case of the glycine receptor. This is because in the open conformation (PDB ID: 3JAD) the pore is dilated so much that there is a significant amount of space between two neighbouring M2 helices. Through these gaps, lipid tails can freely access the pore. A similar behaviour could be observed in the desensitised conformations (PDB IDs: 3JAF and 5VDH), where the distance between the M2 helices increases significantly towards the extracellular end of the TMD. Consequently, productive simulations were in this case carried out for the entire glycine receptor TMD. Importantly, through removing only the ECD, which has no direct role in gating, the size of the simulation system can still be reduced by a factor of $\sim$2 (Figure 4.10A).

As can be seen in Figure 4.10B, the removal of the ECD has no marked influence on pore hydration. In both structures of the closed and desensitised conformations, only either the leucine or the proline gate exhibits dewetting respectively, while in the open conformation the entire channel pore remains fully hydrated. The water density profiles also largely agree quantitatively, with the exception of one closed state structure (PDB ID: 5CFB), where the average density near the leucine gate is slightly lower in simulations

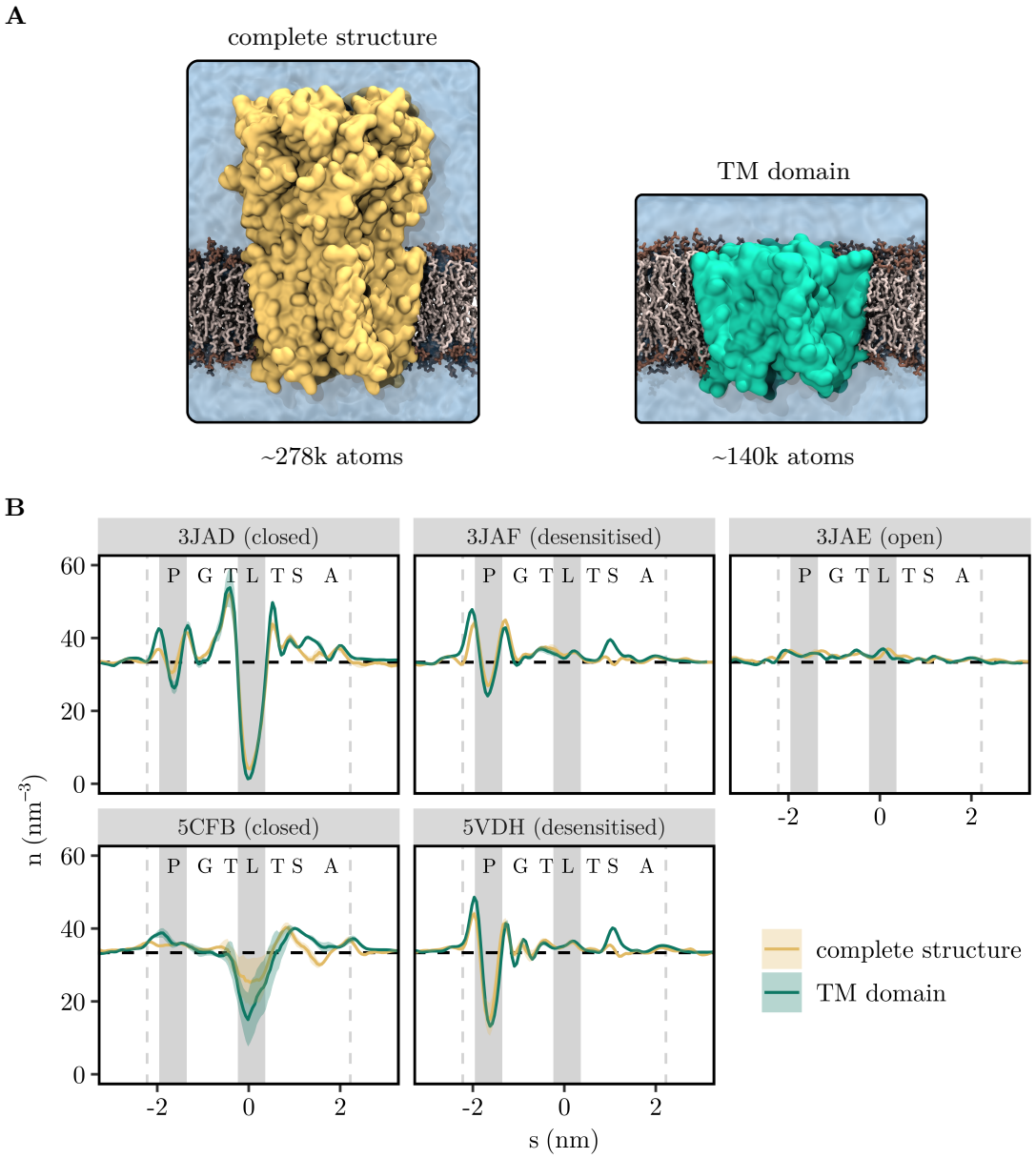



**Figure 4.10:** Influence of glycine receptor structural model. (A) Simulation system based on the complete glycine receptor structure compared to a reduced simulation system, in which the protein is represented by the transmembrane domain only. The latter system comprises a factor of ∼2 fewer particles. Structures shown represent the closed (PDB ID: 3JAD) conformation of the channel. (B) Comparison of time-averaged water density profiles in the transmembrane domain of four different conformations of the glycine receptor. Vertical dashed lines denote the extent of the transmembrane pore and the shaded background indicates the location of the proline and leucine gates. The dashed horizontal line represents the density of bulk water of 33.37 $\mathrm{nm}^{-3}$. Confidence bands represent the standard deviation over three independent MD simulations.

of the TMD. However, this difference is relatively small compared to the variation across repeats. Overall, simulations of the TMD would lead to the same functional annotation as simulations employing the full channel structure.

### 4.2.3 Prevalence of Hydrophobic Gates in Biological Ion Channels

In the previous two sections, it was shown how CHAP can be used to functionally annotate ion channel structures. However, the examples considered so far concerned structurally similar and functionally related members of the pLGIC family. While individual studies have reported hydrophobic gates in a number of unrelated channels (see Section 1.2.2 for an overview), the question remains how widespread hydrophobic gating is and under what conditions it may occur in biological ion channels.

To address this problem, CHAP was employed to determine the average water density in the pores of ∼200 ion channels, representing a substantial fraction of the atomistically resolved ion channel structures currently available in the PDB[§]. This analysis is based on three repeats of 30 ns long MD simulations for each of these structures, amounting to a cumulative simulation time of ∼18 µs. From this population level data, it is possible to derive a phase diagram for the liquid-vapour equilibrium of water confined in biological pores. This is shown in Figure 4.11, where the average density of water is depicted in dependence of pore radius and hydrophobicity.

As can be seen, very narrow pores with radii below ∼0.15 nm typically dewet completely, which is unsurprising as they are too small to accommodate even a single water molecule. However, in intermediate size channels with radii between ∼0.15 nm and ∼0.40 nm, the hydrophobicity of the pore has a strong influence on its hydration behaviour and hydrophobic pores ($H > 0.0$ a.u.) tend to dewet, while hydrophilic pores

[§]S. Rao curated the ion channel structure data set and conducted the MD simulations presented in this study. Data analysis was performed jointly by S. Rao and G. Klesse. A method for calculating pore hydrophobicity profiles was developed by G. Klesse and was implemented in CHAP specifically to facilitate this investigation. Please see the "author contributions" of Rao et al. [103] for further details.

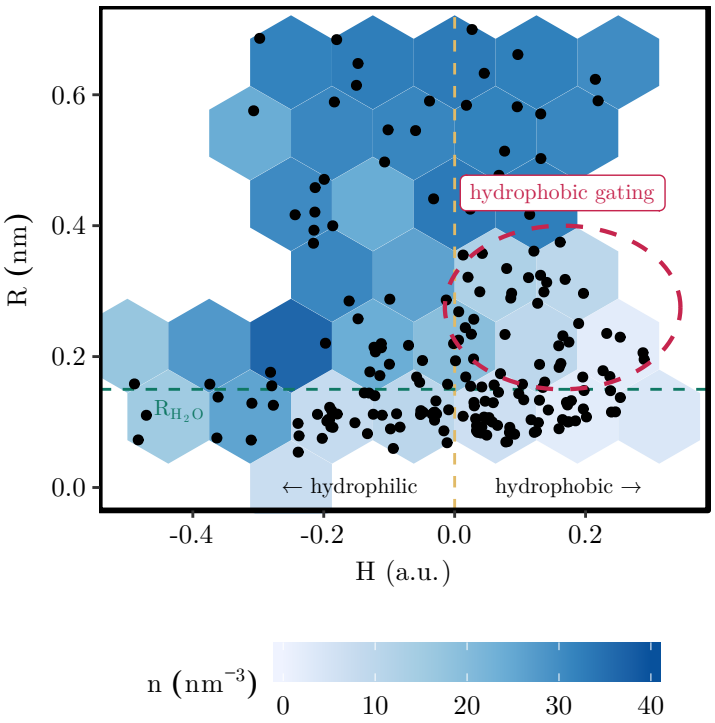



**Figure 4.11:** Pore dewetting across the ion channel structural proteome. Each point corresponds to an individual channel structure and marks the radius and hydrophobicity of its pore at the putative location of the channel gate. Channels are grouped in hexagonal bins, which are coloured by the density of water in the channel pore, averaged over all channels in a bin. The yellow dashed line separates the dataset in predominantly hydrophilic (left) and predominantly hydrophobic (right) pores, while the dashed green line indicates the approximate radius of a water molecule of 0.15 nm. Below this line, water molecules cannot be sterically accommodated in the channel pore. A red circle highlights the hydrophobic gating regime ($H > 0.0$ a.u. and $0.15\,\text{nm} \lesssim R \lesssim 0.40\,\text{nm}$).

($H < 0.0$ a.u.) remain water-filled. In even larger pores, no dewetting is observed and pore water density is close to the density of bulk water of $\sim$33.37 nm$^{-3}$.

Importantly, the hydrophobic gating regime includes pores that could sterically accommodate even fully hydrated ions. For example, the hydration shell of $Na^+$ has a radius of 0.24 nm, while for $Cl^-$ it extends to 0.38 nm. For these channels, functional annotation based purely on the radius of the pore cannot be accurate and hydrophobicity needs to be taken into account explicitly. As can be seen in Figure 4.11, approximately 20 % of the $\sim$200 structures included in the data set belong to this region of the phase diagram. This highlights the general importance of the hydrophobic gating mechanism beyond the particular examples considered previously.

Some limitations of Figure 4.11 should also be noted. Firstly, the theory of hydrophobic gating suggests that the length of the pore (or rather the length of the hydrophobic pore segment constituting a hydrophobic gate) should also impact water density alongside pore radius and hydrophobicity (see Section 1.2.4 for details). However, the ∼200 data points available would not be sufficient to fully sample this three dimensional space. Fortuitously, the variation in the length of hydrophobic pore segments is relatively limited and typically hydrophobic gates are formed by rings of between one (as in the glycine receptor) and three (as in the 5-$HT_3$ receptor) pore lining residues, i.e. they are not normally longer than three helix turns.

Secondly, Figure 4.11 suggests that the average density of water inside moderately hydrophilic ($-0.2\,\text{a.u.} \lesssim H < 0.0\,\text{a.u.}$), intermediate size ($0.15\,\text{nm} \lesssim R \lesssim 0.4\,\text{nm}$) pores lies above the water density in more hydrophobic channels of a similar radius, but below the density of bulk water. This may be due to the fact that the scale of Wimley and White [238] used by CHAP to calculate pore hydrophobicity only approximates the actual conditions inside the ion channel pore, as for example the exact orientation of sidechains is not taken into account. However, in Rao et al. [103] it is shown that using any of the alternative hydrophobicity scales discussed in Section 3.2.3 does not substantially change the shape of the phase diagram and only influences the critical hydrophobicity threshold above which significant dewetting can be observed.

## 4.3 Discussion

In Sections 4.2.1 and 4.2.2, MD simulations of water were used to assign a functional status to several experimentally determined structures of two pLGICs, the 5-$HT_3$ and glycine receptors, which were thus classified as either conducting or nonconducting conformations. How reliable is this functional annotation when compared to more advanced computational techniques? Trick et al. [78] carried out umbrella sampling simulations to derive free energy profiles for $Na^+$ and $Cl^-$ ions in the closed state (PDB ID: 4PIR) struc-

ture of the 5-$HT_3$ receptor, finding high free energy barriers for both ionic species. In addition, through computational electrophysiology simulations, Trick et al. [78] validated that the hydrophobic gate in this structure remains dewetted even under a superphysiological electric field.

Moreover, the functional annotation for the 5-$HT_3$ receptor structures is also in agreement with the results of independent investigations. Through microsecond-long equilibrium MD combined with metadynamics simulations, Yuan et al. [226] concluded that the 4PIR structure represents a closed conformation of the 5-$HT_3$ receptor, which can be opened through rotamer changes of the hydrophobic gate residues (L9′, V13′, and I17′) along with a dilation of the transmembrane pore. Furthermore, in simulations of the desensitised (PDB ID: 6DG7) and open (PDB ID: 6DG8) state structures Basak et al. [225] found hydration patterns for both the transmembrane and intracellular pore comparable to what was presented in Section 4.2.1.

In the case of the glycine receptor, the results presented in Section 4.2.2 are similarly supported by independent simulation studies. Gonzalez-Gutierrez et al. [260] demonstrated that the open conformation structure (PDB ID: 3JAE) is conductive by showing that $Cl^-$ ions can readily permeate the channel if an external electric field is applied. Furthermore, high rates of $Cl^-$ conduction through the 3JAE structure were also observed by Cerdan et al. [261], who additionally demonstrated that the desensitised (PDB ID: 3JAF) structure is impermeable to $Cl^-$ ions.

However, the glycine receptor also illustrates an important limitation of the functional annotation protocol employed here. Based on the results presented in this chapter, no conclusions can be drawn about whether an experimentally determined structure is physiologically relevant. For example, it has been pointed out that the pore of the 3JAE structure of the glycine receptor is ∼0.2 nm wider than the pore of its open state bacterial homologues and in MD simulations this structure exhibits a rate of $Cl^-$ permeation that significantly exceeds the conductance measured in electrophysiology experiments.[260,261]

Consequently, the 3JAE structure has been suggested to be physiologically insignificant and recently refined structures of the glycine receptor open state derived from sophisticated MD simulations (without the positional restraints used in this work) have been proposed.[261,262] Nevertheless, while an analysis of water behaviour cannot elucidate the physiological relevance of a channel structure, it could be used to provide an indication of the conductive status of such an MD-refined structure.

In Section 4.2.3, results of a high throughput analysis of ∼200 ion channels were presented and it was shown that hydrophobic gating is a potentially important mechanism in ∼20 % of all currently available ion channel structures. One limitation of this analysis is that the structures available in the PDB do not provide an unbiased sample of the different functional states of ion channels, as the closed conformation is typically more accessible for the techniques of experimental structural biology than the open state. In Figure 4.11, the preferential sampling of closed states is reflected in a much larger number of small radius pores compared to intermediate sized and large pores. However, the undersampling of open states may conceivably be improved with advances in cryo-EM, which allows multiple conformations to be analysed within a single mixed-population sample. While the phase diagram shown in Figure 4.11 will be refined with the emergence of additional structures, the current sampling is sufficient to support the general conclusion that pore hydrophobicity needs to be taken into account when assigning a functional status to an ion channel structure.

It is also worth noting that Rao et al. [103] developed a machine learning model to functionally annotate ion channel structures without the need for MD simulation. To this end, a support vector machine was trained on the data set presented in Section 4.2.3 to predict the hydration state of an ion channel pore based on its radius and the hydrophobicity of its pore-lining residues. Notably, it was shown by means of ROC curve analysis that a model trained on the basis of CHAP data yields a better performance than a similar model trained solely on HOLE profiles, i.e. without considering hydrophobicity.

This result further emphasises value added by CHAP to the functional annotation of novel ion channel structures.

## 4.4 Summary and Conclusions

The present chapter has provided several examples for the utility of CHAP. Equilibrium MD simulations of water were used to annotate several structures of the 5-$HT_3$ receptor, including its first X-ray structure (PDB ID: 4PIR), the functional status of which had previously been elusive. Subsequently, CHAP was applied to five structures of the glycine receptor and it was demonstrated that water permeation can be used as a reliable proxy for the functional status of these structures as well.

Simulations of the 5-$HT_3$ and glycine receptors were also used to probe the robustness of the simulation protocol first introduced in Section 3.3.1. Firstly, it was shown that harmonic positional restraints can be used to maintain a channel in a conformation close to an experimental structure, without significantly impacting the liquid-vapour equilibrium in its pore. Secondly, it was demonstrated that simulations of a reduced subset of an ion channel structure yield equivalent results to simulations of a complete channel structure, provided that lipid tails are prevented from entering the pore lumen. Such reduced model systems permit significant savings in simulation time, a fact that will be exploited in simulations employing polarisable force fields presented in the following chapter.

Finally, CHAP has enabled a large scale survey of potential hydrophobic gates across the ion channel structural proteome. It was shown that pore dewetting arises in $\sim$20 % of the currently available ion channel structures and frequently occurs in channel pores that are larger than the radii of biologically relevant hydrated ion species. These results highlight the broad relevance of the hydrophobic gating mechanism and underscore the utility of CHAP as a new tool able to detect such gates in future structural biology studies.

Water is a very good servant,
but it is a cruell maister.

*(William Bullein)*

# 5

# The Influence of Induced Polarisation

The results presented thus far are all derived from classical MD simulations employing pairwise additive force fields, in which electrostatic interactions are modelled as Coulomb forces between fixed point charges on atomic centres. However, force fields of this type cannot capture the important physical effect of induced polarisation: depending on the distribution of charges in the environment of an atom, its electronic structure will be distorted compared to that around an isolated atom in a vacuum.

While additive force fields can account for induced polarisation in a mean-field way, they cannot provide a parameterisation that accurately describes a molecule in both the gas and condensed phases. The magnitude of this effect is well exemplified by the water molecule, whose liquid state dipole moment of 2.95 D as obtained from ab initio calculations[283] represents an increase of nearly 60 % over its experimentally measured[284] gas phase value of 1.85 D. A similar difference has been observed for water molecules in the vicinity of a solute protein[285] and the dipole moment of water is also expected to vary depending on whether a solute surface is hydrophobic or hydrophilic.

As a consequence of this, fixed point charge models frequently struggle to accurately reproduce the phase diagram of real water[289] and their ability to make predictions under thermodynamic conditions for which they were not parameterised is likely limited.[130]

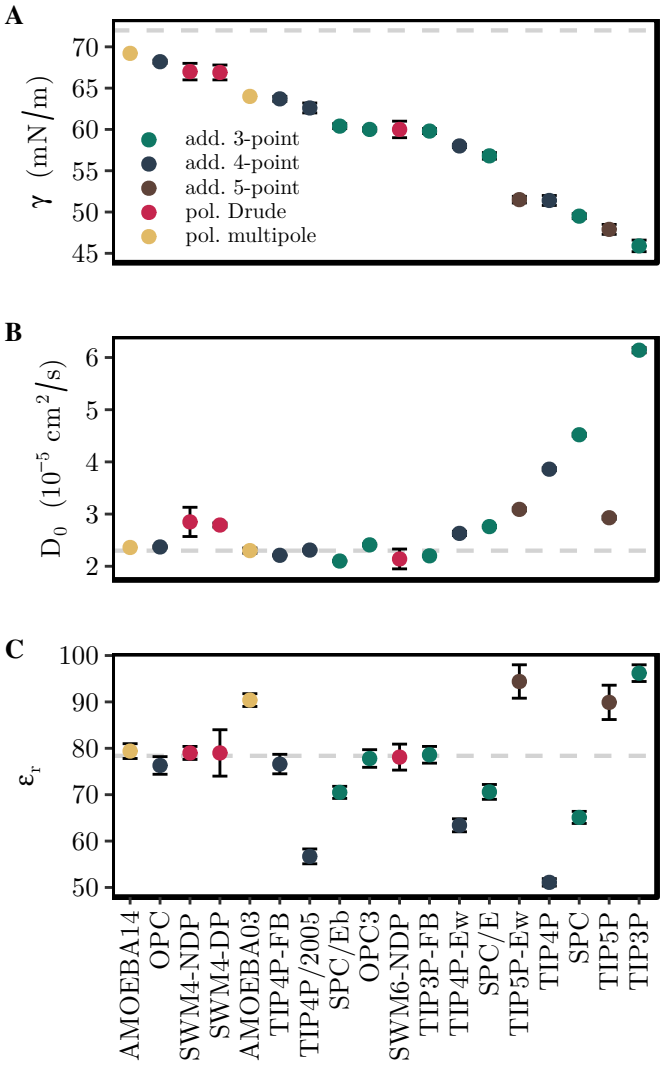



**Figure 5.1:** Macroscopic properties of common water models. (A) Surface tension at the liquid-vapour interface. (B) Self-diffusion coefficient of pure water. (C) Relative electric permittivity of liquid water. Properties of the additive water models were taken from a recent benchmark study[131] and were derived from simulations at 298.15 K and 1 bar. Drude oscillator model properties were extracted from their original publications[165–167] and were determined at the same temperature and pressure. Properties of the polarisable multipole models were taken from a comparison of the AMOEBA03 and AMOEBA14 water models,[186] where simulations were conducted at a pressure of 1 bar and at temperatures of 300 K and 298 K respectively. Where shown, error bars reflect the standard deviation given in the respective publication. Dashed lines represent the experimental reference value of 71.99 mN m$^{-1}$ for the liquid-vapour surface tension,[286] $2.299 \times 10^{-5}$ cm$^2$/s for the self-diffusion coefficient,[287] and 78.375 for the relative permittivity.[288] Water models are ordered according to their accuracy in describing surface tension.

These limitations are particularly relevant in the study of hydrophobic gating, since the equilibrium between the "wet" and "dry" states of a channel pore is controlled by the balance of surface tensions at the liquid-vapour and liquid-protein interfaces.[44] As illustrated in Figure 5.1, the host water models of all common biomolecular force fields (TIP3P, SPC/E, and TIP4P) significantly underestimate the liquid-vapour surface tension of pure water,[129,131,290] which is a central parameter in the thermodynamic theory of hydrophobic gating (Section 1.2.4).

This deficiency of the pairwise additive approximation has long been recognised[291,292] and over the last two decades polarisable force fields have been developed for water[293] as well as for biomolecular systems.[155,294] Early efforts include the AMBER polarisable force field[295] and the charge equilibration force field (CHEQ),[296] but perhaps the most prominent approaches are the CHARMM Drude[160] and AMOEBA[99] force fields. They are based on two different theoretical approaches to polarisation in a classical MD framework, Drude oscillators[161] and polarisable atomic multipoles,[183] and provide consistent parameter sets not only for proteins,[168,190] but also for membrane lipids.[171,192] The theoretical background of these force fields is discussed in Chapter 2.

A notable success of polarisable force fields is their ability to faithfully reproduce the properties of bulk water. As can be seen in Figure 5.1, out of the five models that best reproduce the surface tension at the liquid-vapour interface, four (the AMOEBA03,[185] AMOEBA14,[186] SWM4-DP,[165] and SWM4-NDP[166] models) are polarisable and only one (the OPC[124] model) is a fixed point charge model. Importantly, the good performance of these models is not limited to surface tension and other important quantities like the self-diffusivity and the electric permittivity are in quantitative agreement with experiment as well. This indicates that polarisable models correctly account for the underlying physics and suggests that they are transferable.

Indeed, the AMOEBA03 model reproduces many thermodynamic properties of water over a wide range of temperatures and pressures despite only being parameterised against bulk properties at ambient conditions.[297] Its phase behaviour has also been shown to agree well with experiment for temperatures of up to 100 K above room temperature.[298] The reparameterised AMOEBA14 model improves upon this and reliably predicts the properties of liquid and gas phase water for a large portion of its phase diagram.[186] Similarly, the Drude oscillator model SWM4-NDP also generalises well to elevated temperatures and pressures, showing comparable performance to the AMOEBA03 model.[299] Both the AMOEBA models[187] and the SWM4-NDP model[300] also accurately describe

the thermodynamics of ion solvation. For example, a recent study of the gramicidin A channel[101] suggests that the explicit inclusion of polarisation effects significantly improves computational predictions of ion channel conductance.

However, the past two decades have also seen the development of a large number of novel additive water models, many of which achieve a significantly improved accuracy compared to the traditional TIP3P, SPC/E, and TIP4P models (Figure 5.1). Of the water models shown in Figure 5.1, TIP4P/2005[123] and OPC[124] are particularly noteworthy. TIP4P/2005 was specifically developed as a general purpose model that accurately describes the phase behaviour of water. Its phase diagram agrees well with experiment[301] and it was found to be more accurate than many other rigid point charge models.[130,289] The OPC water model is based on a novel global optimisation procedure and under ambient conditions it achieves a level of accuracy comparable to that of polarisable water models (Figure 5.1). Its bulk properties agree with experimental data over a wide range of temperatures,[124] but its phase diagram has not been fully mapped yet.

Considering the improved transferability of polarisable force fields and the success of their associated water models in quantitatively describing the relevant properties of the bulk liquid, it is imperative to employ them in the investigation of hydrophobic gates. However, since polarisable MD is significantly more computationally expensive than conventional atomistic MD, it is important to also evaluate the accuracy of modern additive water models. The primary aim of this chapter therefore is to quantify the magnitude of induced polarisation experienced by water molecules passing through an ion channel and the impact this has on the hydration equilibrium of the channel pore. Furthermore, the influence of polarisation effects on the permeation of atomic ions is analysed. This investigation employs both the CHARMM Drude and AMOEBA polarisable force fields as well as variety of fixed point charge water models in conjunction with the additive CHARMM36m protein force field.

The remainder of this chapter is organised as follows. Section 5.1 describes protocols

for polarisable MD simulation of membrane protein systems and provides details of the analyses underlying the results presented in Section 5.2. A validation of these simulation protocols is discussed in Section 5.2.1. In Section 5.2.2 the pore hydration behaviour of polarisable water models is compared to that of fixed point charge models and the dipole moment of water in hydrophobic confinement is quantified. Subsequently, Section 5.2.3 examines the influence of induced polarisation on ion conduction. These results are discussed in Section 5.3 and the findings are summarised in Section 5.4.

## 5.1 Methods

This investigation uses the 5-$HT_3$ receptor discussed in Section 4.2.1 as a test system, as the existence of a hydrophobic gate in this ion channel is well established.[78,226] Moreover, experimental structures are available for different conformational states, facilitating a direct comparison of water models at various degrees of pore constriction. In particular, the PDB structures with IDs 4PIR (apo/closed state, X-ray crystallography),[255] 6BE1 (apo/closed state, cryo-EM),[256] 6DG7 (5-$HT_3$-bound/intermediate, cryo-EM),[225] and 6DG8 (5-$HT_3$-bound/open state, cryo-EM)[225] were considered here.

### 5.1.1 Equilibrium Molecular Dynamics Simulation

As polarisable MD simulations can be very computationally demanding, only the pore-lining M2 helix bundle (residues 247 – 271) embedded in a DOPC bilayer was simulated. To test the impact of this simplification on the induced polarisation of water inside the channel pore, additional simulations based on the 4PIR structure were conducted using either the M2 helix bundle in a POPE bilayer or the full TM domain (residues 220 – 334 as well as 426 – 459) in a DOPC bilayer. Equilibrium MD simulations were performed for all four structures, while umbrella sampling simulations of ion permeation were only carried out for the 6DG8 structure.

To assess the hydration behaviour of different water models in the hydrophobic pore

of the 5-$HT_3$ receptor, a series of equilibrium MD simulations were carried out. Productive simulations using the additive CHARMM36m and polarisable CHARMM Drude force fields had a total duration of 150 ns, while for the more computationally intensive AMOEBA force field, the simulation time was reduced to 50 ns. Three independent repeats were carried out for each unique combination of parameters and the first 10 ns of each simulation were regarded as equilibration period and were not included in the analysis. The details of the respective simulation protocols are described below.

### Structural Model and System Preparation

Ion channel structures were obtained from the PDB and missing atoms were added using the WHAT IF tool.[242] The channel protein was then embedded in a homogeneous lipid bilayer using a serial multiscale procedure.[227] Herein, the protein was first converted to a coarse-grained representation and coarse-grained lipids, ions, and water particles were added to the simulation box. A 100 ns simulation employing the MARTINI force field[302] was then carried out using GROMACS 5.1,[98] leading to the formation of a lipid bilayer around the transmembrane domain of the protein. Throughout this simulation, backbone positional restraints of 1000 kJ/mol/nm$^2$ were applied to prevent large scale conformational changes of the protein.

Following coarse-grained bilayer self-assembly, the protein-bilayer system was converted back to an atomistic representation and resolvated in a NaCl solution with a salt concentration of 0.15 mol $L^{-1}$. This was followed by a 10 ns simulation with the additive CHARMM36m force field,[109] the corresponding lipid parameters,[303] and the CHARMM-modified mTIP3P water model[120] using the same parameters as the productive simulations described in the following section.

This simulation served two purposes: firstly, it allowed water molecules placed in unphysical locations (e.g. between the two leaflets of the lipid bilayer or in isolated molecular cavities in the protein) during the atomistic solvation step to escape to the

bulk water regime (trapped water molecules were removed from the simulation system prior to productive simulation). Secondly, it provided a system configuration in which the atomistic degrees of freedom are equilibrated, which could be used as an initial configuration for polarisable MD. Atomistic equilibration using the CHARMM additive force field is suggested in the literature[168] for simulations with the CHARMM Drude force field, but has also proven helpful to ensure stable dynamics during simulations employing the AMEOBA force field.

#### CHARMM Additive Force Field Simulations

As a basis for comparison, atomistic simulations using the additive CHARMM36m protein force field[109] with associated lipid parameters[303] and the mTIP3P water model[120] were performed. Additionally, the CHARMM36m force field was combined with a number of further fixed point charge water models. A comprehensive comparison of all existing additive water models is beyond the scope of this study. Instead, the SPC/E[122] and TIP4P[118] models were chosen as examples of widely used yet older models and the more recent TIP4P/2005[123] and OPC[124] models were employed due to their superior accuracy.

These simulations were carried out in GROMACS 2018 using a leapfrog integrator with a step size of $\Delta t = 2\,\mathrm{fs}$ for time integration. The length of bonds involving hydrogen atoms was constrained through the LINCS algorithm[139] and water geometry was kept rigid using the SETTLE method.[137] To preserve the experimentally determined protein structure, a harmonic restraining potential of $1000\,\mathrm{kJ/mol/nm^2}$ was applied to all C$\alpha$ atoms.

The smooth PME[152] method with a real space cutoff of 1.0 nm, a Fourier spacing of 0.12 nm, and charge interpolation through fourth order B-splines was used to calculate electrostatic interactions under periodic boundary conditions. Van der Waals interactions were smoothly switched off between 1.0 nm and 1.2 nm and a long range dispersion

correction was applied to both energy and pressure. Simulations were performed in the isothermal-isobaric ensemble at a temperature of 310 K and a pressure of 1 bar. A velocity rescaling thermostat[142] with a coupling constant of $\tau_T = 0.1\,\mathrm{ps}$ and a semi-isotropic Parrinello-Rahman barostat[143] with a coupling constant of $\tau_P = 1\,\mathrm{ps}$ and a compressibility of $4.5 \times 10^{-5}\,\mathrm{bar}^{-1}$ were used for temperature and pressure control respectively.

### CHARMM Drude Force Field Simulations

Starting configurations for simulations using the CHARMM polarisable force field were generated from the atomistically equilibrated lipid-bilayer systems described in Section 5.1.1 by adding Drude oscillators with the DrudePrepper functionality of CHARMM-GUI.[304] Unfortunately, DrudePrepper requires input topologies in either CHARMM or NAMD PSF format and does not accept the TOP format files used by GROMACS. PSF topologies were therefore generated using the `psfgen` utility of VMD 1.9.2[248] and the minimisation stage of the DrudePrepper pipeline was executed in NAMD 2.13.[305] Productive simulations were then carried out in OpenMM 7.3.0 in mixed precision mode.

The Drude 2013 protein force field[168] was used together with Drude polarisable lipid parameters,[171] the SWM4-NDP water model,[166] and compatible ion parameters.[300] Particle positions were advanced using a dual-Langevin integrator with a time step of $\Delta t = 1\,\mathrm{fs}$. Atomic nuclei were coupled to a heat bath at 310 K using a Langevin friction of $5\,\mathrm{ps}^{-1}$, while Drude oscillators were maintained at a temperature of 1 K with a friction coefficient of $20\,\mathrm{ps}^{-1}$. In addition, an anisotropic Monte Carlo barostat was used to enforce a pressure of 1 bar, with trial volume changes attempted every 20 integration steps.

Periodic boundary conditions were used and Coulombic forces were calculated using the PME method[152] with a real space cutoff of 1.0 nm, an error tolerance of $1 \times 10^{-4}$, and charge interpolation onto a regular grid through fifth order B-splines. Van der Waals forces were smoothly switched off over the distance interval from 1.0 nm to 1.2 nm

and long range dispersion interactions were taken into account through a correction to the potential energy.[153] C$\alpha$ atoms were subjected to a harmonic restraining force of 1000 kJ/mol/nm$^2$ to limit deviations from the experimentally determined protein structure. Water molecule geometry was kept rigid and the length of covalent bonds involving hydrogen was constrained. Additionally, a Drude hard-wall constraint[170] was employed to prevent Drude oscillators from moving further than 0.02 nm from their respective nucleus.

### AMOEBA Force Field Simulations

Polarisable atomic multipole simulations were carried out in OpenMM 7.1.1 using the AMOEBA13 protein parameter set[190] in conjunction with both the AMOEBA03[185] and AMOEBA14[186] water models, as well as polarisable multipole parameters for ions[187] and lipids.[192] Starting configurations for these simulations were derived from the final frames of a 10 ns long atomistic equilibration simulation as described in section 5.1.1. However, productive simulations were preceded by 1000 steps of energy minimisation with the AMOEBA force field to avoid divergent energies due to induced dipoles.

Time integration was performed using the r-RESPA method[207] with an outer time step of $\Delta t = 2$ fs and an inner time step of $\delta t = 0.25$ fs. Electrostatic multipole and van der Waals forces were updated in intervals of the outer time step, all other forces were evaluated at each inner time step. OpenMM was operated in mixed precision mode with forces being accumulated in single precision and time integration performed in double precision.

Electrostatic multipole interactions were treated through the PME method,[185,306] with a real space cutoff distance of 0.8 nm, an Ewald error tolerance of $5 \times 10^{-4}$, and interpolation performed through fifth order B-splines. Van der Waals forces were computed up to a distance of 1.2 nm and a long range dispersion correction[153] was applied to account for interactions beyond this cutoff. Induced dipoles were computed through truncated

self-consistency iteration using the extrapolated perturbation theory method.[206]

The simulation system was maintained at a temperature of 310 K using an Andersen thermostat[140] with a collision frequency of $\tau_T = 1\,\mathrm{ps}^{-1}$. An anisotropic Monte Carlo barostat[146] was used to enforce a system pressure of 1 bar, with Monte Carlo moves attempted every 20-th integration step. In accordance with the AMOEBA model, no constraints on covalent bonds or water molecule geometry were enforced. However, in order to restrain the protein in a configuration close to its experimentally determined structure, all C$\alpha$-atoms were placed under a harmonic restraint with a force constant of $1000\,\mathrm{kJ\,mol^{-1}\,nm^{-2}}$.

### 5.1.2 Umbrella Sampling

Umbrella sampling was performed to obtain one-dimensional single ion potential of mean force profiles for both $Na^+$ and $Cl^-$ ions passing through the pore of the open state structure (PDB ID: 6DG8) of the 5-$HT_3$ receptor. These simulations were carried out using the AMOEBA force field with the AMOEBA14 water model, the CHARMM Drude force field with the SWM4-NDP water model, and the CHARMM36m force field with both the mTIP3P and TIP4P/2005 water models. Simulation protocols were similar to those detailed above for equilibrium MD simulations.

The collective variable was chosen to be the distance between the ion and the centre of mass of the protein along the $z$-axis of the simulation box, which corresponds to the direction normal to the lipid bilayer membrane. Starting configurations for the umbrella windows were generated from the final state of the equilibrium MD simulations described in Section 5.1.1. A target ion was relocated to subsequent positions along the $z$-axis and the lateral position of the ion was set to the centre of mass of the protein to ensure a placement inside the channel pore. This was followed by 10 steps of energy minimisation to remove steric clashes between the target ion and water molecules present in the pore. During both minimisation and sampling, a harmonic biasing force of $2000\,\mathrm{kJ\,mol^{-1}\,nm^{-2}}$

was applied to restrain the collective variable.

Umbrella windows covered the entire length of the ion channel and reached up to 1 nm into the bulk water regime, while keeping a minimum distance (in the initial configuration) of 0.5 nm from the boundary of the simulation box to allow for box size variations due to barostat coupling. Each umbrella window was simulated for 10 ns. The distance between two subsequent umbrella windows was 0.1 nm for simulations with the CHARMM36m and AMOEBA force fields, and 0.05 nm for simulations with the CHARMM Drude force field. The former setup corresponds to 62 windows with a cumulative simulation time of 620 ns for each of $Na^+$ and $Cl^-$, while the latter corresponds to 112 windows and a simulation time of 1.12 µs per ion species.

Unbiasing was performed through the weighted histogram analysis method (WHAM) using the Grossfield lab `wham` implementation in version 2.0.9.[211] The first half of each simulation window was regarded as equilibration period and the final PMF profile, $F(z)$, was calculated only from the final 5 ns of simulation time. Convergence was assessed by calculating an additional PMF profile, $F_t(z)$, from a shorter time interval, $[t - \tau/2, t + \tau/2)$, centered around time $t$, where an interval length of $\tau = 2$ ns was used.

This instantaneous PMF profile can be compared to the final profile by means of the mean signed deviation,

$$\mathrm{MSD}(t) = \frac{1}{N_{\mathrm{win}}} \sum_{i=1}^{N_{\mathrm{win}}} \left(F_t(z_i) - F(z_i)\right), \tag{5.1}$$

and the root mean squared deviation,

$$\mathrm{RMSD}(t) = \sqrt{\frac{1}{N_{\mathrm{win}}} \sum_{i=1}^{N_{\mathrm{win}}} \left(F_t(z_i) - F(z_i)\right)^2}, \tag{5.2}$$

where $z_i$ represents the value of the collective variable of the $i$-th window and $N_{\mathrm{win}}$ is the total number of windows. For a converged PMF, the value of these error metrics tends to zero with increasing simulation time, provided the interval length is large enough.

### 5.1.3 Analysis

#### Water Density and Beckstein Openness

A time-dependent water density profile, $n(s,t)$, along the channel pore was calculated using CHAP 0.9.1.[26] From this, the Beckstein openness,[44] a measure of the pore hydration probability discussed in Section 1.2.4, was derived according to

$$\omega(t) = \begin{cases} 1 \text{ if } \min_s n(s,t) > n_{\text{thres}} \\ 0 \text{ otherwise} \end{cases} \tag{5.3}$$

where $n_{\text{thres}} = 7.6\,\text{nm}^{-3}$ is a density threshold used to distinguish between the open/liquid ($\omega(t) = 1$) and closed/vapour ($\omega(t) = 0$) states of the pore. The time-averaged openness,

$$\langle\omega\rangle = \frac{1}{T}\int_0^T \omega(t)\ dt, \tag{5.4}$$

was then used to estimate the free energy difference,

$$\Delta\Omega = \Omega_{\text{vapour}} - \Omega_{\text{liquid}} = -\ln\left(\frac{1-\langle\omega\rangle}{\langle\omega\rangle}\right), \tag{5.5}$$

between these two states. Note that when $\langle\omega\rangle \in \{0,1\}$ simulations were too short to observe transitions between the liquid and vapour states and consequently the free energy difference cannot be estimated.

#### Water Molecular Dipole Moment

To quantify the effect of electronic polarisation experienced by water passing through a hydrophobic gate, the magnitude and orientation of its molecular dipole moment was determined from equilibrium MD simulations. For the water models considered here, the dipole moment of an individual water molecule comprises three contributions corresponding to spatially distributed charge monopoles, $\mathbf{M}^{\text{mon}}$, induced dipoles, $\mathbf{M}^{\text{ind}}$,

and permanent atomic dipoles, $\mathbf{M}^{\mathrm{per}}$. Since water molecules carry no net charge, the monopole contribution is unambiguously defined by

$$\mathbf{M}^{\mathrm{mon}} = \sum_{i \in \mathcal{I}} q_i \mathbf{r}_i, \tag{5.6}$$

where $\mathcal{I} = \{\mathrm{O}, \mathrm{H}_1, \mathrm{H}_2\}$ for the three-site additive and AMOEBA models and $\mathcal{I} = \{\mathrm{M}, \mathrm{H}_1, \mathrm{H}_2\}$ for the four-site additive and Drude models. The induced dipole contribution is given by

$$\mathbf{M}^{\mathrm{ind}} = \sum_{i \in \mathcal{I}} \boldsymbol{\mu}_i^{\mathrm{ind}} \tag{5.7}$$

for the AMOEBA models, where $\boldsymbol{\mu}_i^{\mathrm{ind}}$ denotes the induced atomic dipole moment, and by

$$\mathbf{M}^{\mathrm{ind}} = \sum_{j \in \mathcal{J}} q_j \mathbf{r}_j \tag{5.8}$$

for the CHARMM Drude force field, where $\mathcal{J} = \{\mathrm{O}, \mathrm{D}\}$. It is zero for all additive models. Finally, the permanent dipole contribution is nonzero only for the AMOEBA models, for which it is given by

$$\mathbf{M}^{\mathrm{per}} = \sum_{i \in \mathcal{I}} \boldsymbol{\mu}_i^{\mathrm{per}}, \tag{5.9}$$

where $\boldsymbol{\mu}_i^{\mathrm{per}}$ refers to the permanent atomic dipole moment.

A one-dimensional profile along the channel centre line was calculated for the magnitude of the molecular dipole moment of water according to

$$M(s) = \frac{1}{T} \int_0^T \frac{1}{N_s(t)} \sum_k \|\mathbf{M}_k\| \, dt, \tag{5.10}$$

where $\mathbf{M}_k = \mathbf{M}_k^{\mathrm{mon}} + \mathbf{M}_k^{\mathrm{ind}} + \mathbf{M}_k^{\mathrm{per}}$ denotes the total dipole moment of the $k$-th water

molecule. The summation over $k$ includes all water molecules, which at time $t$ are located in a bin centered at $s$ and $N_s(t)$ denotes the overall number of water molecules in the bin at this time. Water molecules were assigned to bins based on the position of their centre of mass. Since the AMOEBA water models are flexible, the magnitude of the monopole and permanent dipole terms may be time- and environment-dependent. Profiles according to Equation 5.10 were therefore also determined for each individual contribution.

Analogously, a profile for the orientational order of the molecular dipole moment was calculated according to

$$S_1(s) = \frac{1}{T} \int_0^T \frac{1}{N_s(t)} \sum_k \cos\theta_k \, dt, \tag{5.11}$$

where $\theta_k$ denotes the angle between the molecular dipole vector and the Cartesian $z$-axis.

Dipole profile calculations were performed using custom scripts based on version 0.19.0 of the MDAnalysis library.[307] CHAP 0.9.1 was used to determine a pore centre line, which was discretised into equidistant bins of 0.1 nm length. Lab frame induced and permanent atomic dipole moments for the AMOEBA models were calculated using OpenMM 7.1.1 by rerunning the truncated self-consistency iteration on frames extracted from the equilibrium MD trajectories.

### Ion Hydration Number

First and second shell hydration numbers of $Na^+$ and $Cl^-$ ions along the channel pore were determined as follows. The radial distribution function (RDF) of water oxygens around the restrained ion was computed individually for each umbrella window according to

$$g_{\text{I–O}}(r) = \frac{1}{N_\text{O}} \sum_{i=1}^{N_\text{O}} \langle \delta(|\mathbf{r}_\text{I} - \mathbf{r}_i| - r) \rangle \tag{5.12}$$

where $I \in \{\mathrm{Na}, \mathrm{Cl}\}$ denotes the species of the ion and $N_\mathrm{O}$ is the number of water molecules inside the reference volume, $V$, of the RDF calculation.

From the RDF, the hydration number was determined as

$$N_\mathrm{I}^i = \frac{4\pi N_\mathrm{O}}{V} \int_{R_{i-1}}^{R_i} g_{\mathrm{I–O}}(r)\, r^2\, dr \tag{5.13}$$

where $R_i$ denotes the limiting radius of the $i$-th hydration shell. The limiting radii for the first, $R_1$, and second, $R_2$, shell were determined from the RDF of ions in the bulk water regime outside the channel as the local minima of the RDF after its first and second peak respectively and $R_0 = 0$. Higher order peaks could not be clearly identified in the RDF.

MDAnalysis 0.19.0 was used to calculate RDFs using 1500 bins up to a maximum distance of 1.5 nm. For simulations with additive force fields, RDF calculations were repeated using GROMACS' built-in tools and the resulting RDF was found to be identical to that computed with MDAnalysis. As a further check, hydration numbers found via RDF integration were validated by directly counting the number of water oxygens within the shells defined by $R_i$.

## 5.2 Results

The behaviour of polarisable water in the hydrophobic gate of the 5-$HT_3$ receptor is examined in Section 5.2.2. Findings on how the inclusion of polarisation effects modifies the free energy barrier for ion permeation in the 5-$HT_3$ receptor are then presented in Section 5.2.3. Before turning to these key results, the following section will demonstrate the validity of the protocols for simulating membrane proteins using polarisable force fields described in Section 5.1. A comparison of the computational cost associated with employing polarisable force fields is given in Appendix A.

### 5.2.1 Validation of Polarisable Simulation Protocol

Parameters for simulations with the CHARMM Drude and AMOEBA force fields were selected in accordance with the conditions under which the respective force field was originally parameterised. However, not many studies that employ polarisable force fields to investigate membrane proteins have been conducted thus far. It was therefore important to ascertain that the chosen parameters permit stable integration and sampling from the appropriate thermodynamic ensemble. In addition, benchmark data was required to determine the feasibility of polarisable simulations on biologically relevant systems. Thus, exploratory simulations were performed using the M2 helix bundle from the closed state structure (PDB ID: 4PIR) of the 5-$HT_3$ receptor embedded in a DOPC lipid bilayer as a test system.

#### AMOEBA Force Field

For polarisable multipole simulations, the inner and outer time step of the r-RESPA integrator can be chosen independently. Preliminary simulations showed that stable integration was impossible with an inner time step of more than 1 fs, which is unsurprising given that hydrogen bonds and water model geometry are unconstrained in the AMOEBA model. The maximum stable outer time step was found to be 4 fs. These limiting time steps were found to be consistent between simulations using the extrapolated perturbation theory method and those employing a full self-consistency iteration to calculate induced dipole moments. As the extrapolation method achieved simulation speeds of up to 50 % faster than the full iteration, all subsequent simulations employed this approach.

Figure 5.2 shows how the thermodynamic properties* of the exploratory simulation system vary for different values of the outer time step. It can be seen that the system

*All simulations were performed in the NPT ensemble. However, the Monte Carlo barostat implementation in OpenMM never calculates the pressure itself and hence box volume is shown instead.

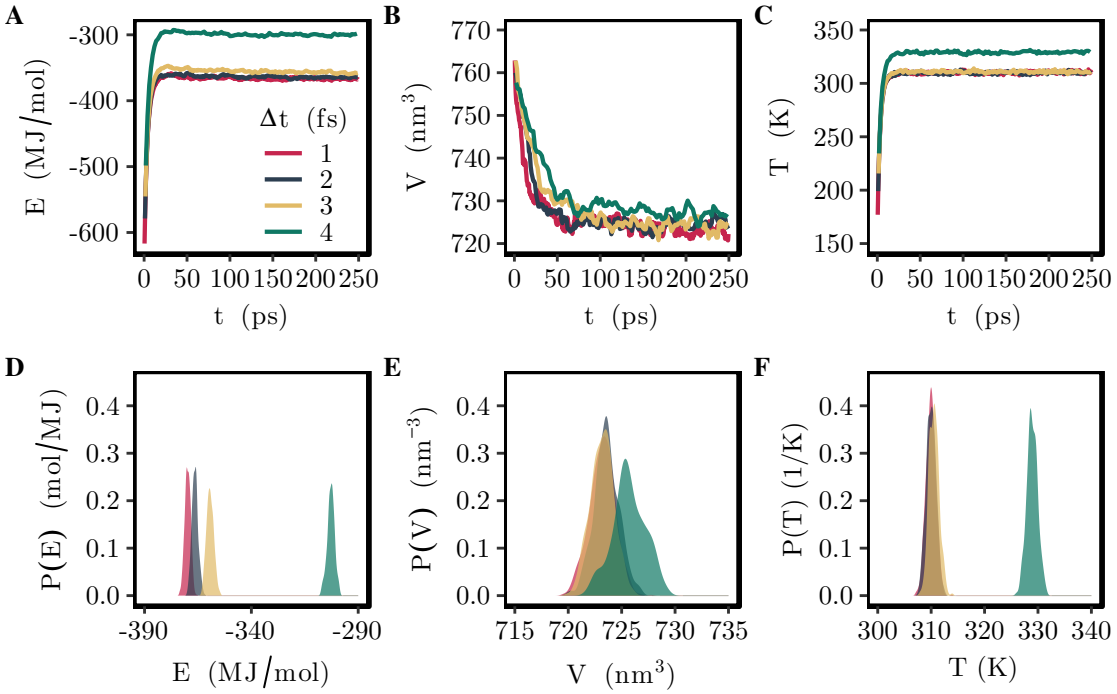



**Figure 5.2:** Equilibration of the polarisable multipole simulation system. (A – C) Time series of total energy, $E$, box volume, $V$, and average temperature, $T$, over the initial 250 ps of a 1 ns long simulation with the AMOEBA force field. All three quantities converge to their equilibrium value within approximately 100 ps. (D – F) Probability densities estimated from samples taken from the equilibrated part of the trajectories ($t > 250$ ps). For an integrator time step of more than 2 fs, the thermodynamic properties of the system are inconsistent with those calculated using a more conservative time step. Data shown was derived from simulations employing the AMOEBA03 water model.

reached its equilibrium quickly, within much less then 1 ns. This fast convergence is observed irrespective of time step. However, for an outer time step of more than 2 fs, the ensemble properties converge to different values from those reached with smaller time steps. For $\Delta t = 3$ fs, only the total energy is affected, but for $\Delta t = 4$ fs, all ensemble properties clearly deviate from their target values, the temperature even by as much as 20 K. Based on these observations, an outer and inner time step of 2 fs and 0.25 fs respectively were adopted for productive simulations. Note that this is slightly more conservative than the time step of 2.5 fs used in the parameterisation of the AMOEBA protein force field,[190] which was performed using soluble rather than membrane proteins.

### CHARMM Drude Force Field

For Drude oscillator simulations, preliminary trials showed that stable integration was possible for time steps <1.5 fs. For larger time steps, the total energy diverges on the picosecond time scale. Figure 5.3 displays how total energy, box volume and average temperature depend on the choice of time step within the stable regime. It can be seen that all properties converge to their respective equilibrium value, which is not significantly altered by small changes of time step size. However, while box volume reaches equilibrium within ∼1 ns, total energy approaches its limiting value much more slowly. In addition, total energy decreases even though the simulation was preceded by local energy minimisation. This indicates that some large scale motion takes place in the system.

Figure 5.4 compares the protein-bilayer system at the end of a 10 ns long equilibration period using the polarisable CHARMM Drude and AMOEBA force fields as well as the CHARMM36m additive force field. As can be seen in Figure 5.4B, Drude oscillator equilibration leads to a protein whose position is shifted towards the intracellular domain. Moreover the lipid bilayer is distorted compared to the planar bilayer resulting from equilibration using the CHARMM36m additive force field (5.4A). It is this rearrangement of the lipid molecules in the vicinity of the protein structure that causes the decrease of internal energy observed in Figure 5.3A.

Given that equilibration with the AMOEBA force field leads to a planar bilayer similar to that obtained with the CHARMM additive force field (cp. Figure 5.4C), it is unlikely that the lipid rearrangement is a consequence of electronic polarisation. Instead, it probably represents an artefact of the lipid parameterisation of the CHARMM Drude force field. Since the M2 helix bundle is a reduced model of the full 5-$HT_3$ receptor that would not normally be in direct contact with the bilayer, it is not sensible to attribute physiological significance to the observed differences in protein positioning.

It is, however, important to check that lipids do not interfere with the intended sim-

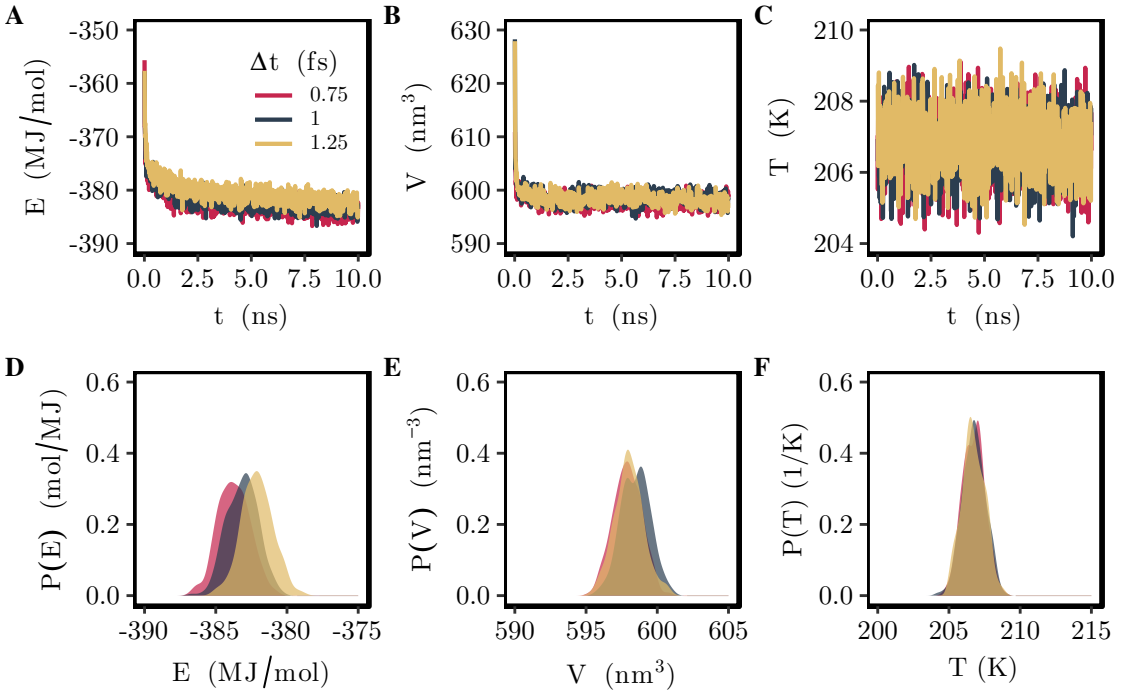



**Figure 5.3:** Equilibration of Drude oscillator simulation system. (A – C) Time series of total energy, E, box volume, V, and average temperature, T, over the course of a 10 ns long simulation with the CHARMM Drude force field. Due to the use of a stochastic Langevin thermostat, temperature is close to its equilibrium value from the start of the simulation. Box volume rapidly converges towards its equilibrium value within less than 1 ns. However, total energy converges more slowly and takes over 5 ns to reach equilibrium. (D – F) Probability densities estimated from samples taken from the equilibrated part of the trajectory ($t > 7.5$ ns). Thermodynamic ensemble properties are consistent for small changes of the integrator time step. Note that the temperature shown represents an average over Drude oscillators and nuclei and does not correspond to the physical target temperature of 310 K.

ulations of water in the receptor pore. As can be seen in Figure 5.4E, lipid headgroups can get closer to the extracellular opening of the channel in the Drude oscillator system, but they do not obstruct the pore itself. Since the hydrophobic gate is located in the interior of the channel, the influence of the observed bilayer deformation on pore hydration is therefore likely to be negligible. However, the process of lipid rearrangement and the associated time scale of ∼5 ns need to be taken into account in the equilibration protocol.

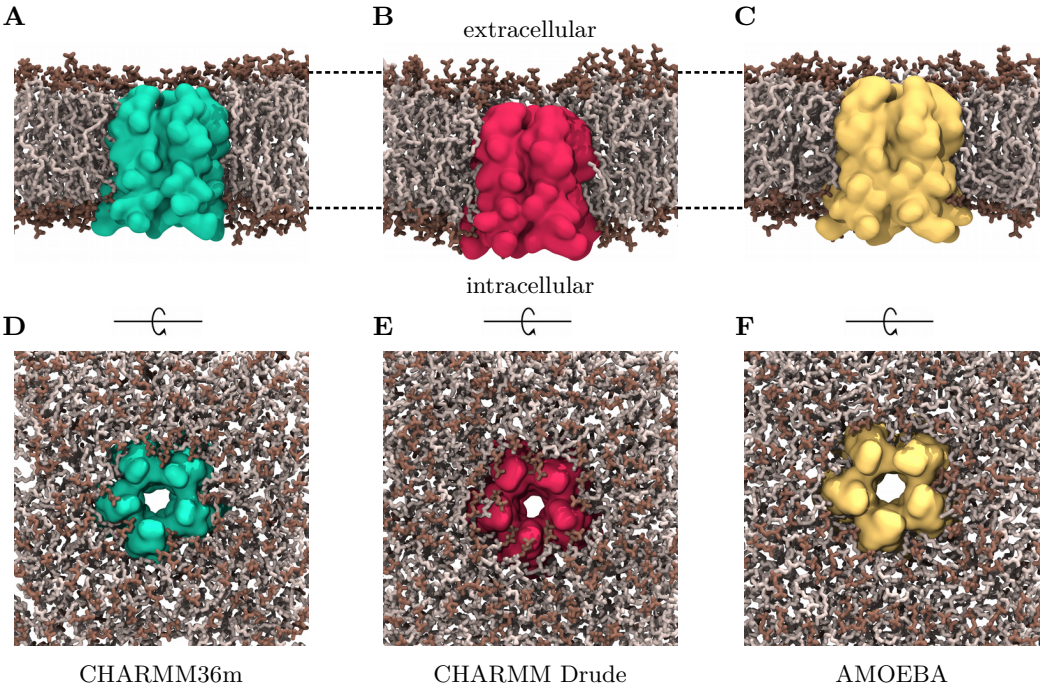



**Figure 5.4:** Protein position relative to lipid bilayer after equilibration with different force fields. Shown is a surface representation of the M2 helix bundle of the 5-$HT_3$-receptor structure 6DG8 embedded in a DOPC bilayer as viewed from the side (A, B, C) and from the extracellular domain (D, E, F). The protein-bilayer system is shown at the end of a 10 ns long simulation using the additive CHARMM36m force field (A, D), the polarisable CHARMM Drude oscillator force field (B, E), and the AMOEBA polarisable multipole force field (C, F). Polarisable simulations were initialised with coordinates resulting from equilibration with the CHARMM36m force field. Lipids are displayed in licorice representation with headgroups and hydrocarbon tails coloured in dark and light shades of brown respectively. Water molecules, ions, and lipid hydrogens are omitted for clarity and lipid molecules obstructing the view of the protein are not shown in panels A, B, and C. In the CHARMM oscillator force field, the bilayer is more strongly deformed compared to the other two cases and the protein is positioned slightly lower in the membrane.

### 5.2.2 Influence of Polarisation on Water Behaviour

#### Pore Hydration

Figure 5.5 compares the pore hydration behaviour predicted by polarisable MD simulations with that observed when employing the additive CHARMM36m force field combined with the mTIP3P water model. As was discussed in Chapter 4, additive MD simulations predict a complete de-wetting of the hydrophobic gate in both the closed (PDB ID: 4PIR and 6BE1) and intermediate (PDB ID: 6DG7) state structures of the

5-$HT_3$ receptor. In Figure 5.5A, it can be seen that for these structures the hydrophobic region of the pore is consistently predicted to be fully de-wetted by all three polarisable water models, in good agreement with the additive result. In the remainder of the channel pore, the predicted water density is close to the density of bulk water and there is again a high level of agreement between the different water models.

The time evolution of the phase state of water in the hydrophobic gate is shown in Figure 5.5B. In the closed and intermediate state structures, water in the hydrophobic gate remains in a vapour state throughout most of the simulation, with only a few brief wetting events occurring in some repeats. This behaviour is remarkably consistent across all four water models and indicates that in these conformational states of the channel, the free energy cost of hydrating the pore is particularly high. While this is consistent with the expectation of a closed hydrophobic gate (at least for the 4PIR and 6BE1 structures), it renders a quantitative comparison of water models infeasible. In order to obtain accurate estimates of the equilibrium constant for transitions between the liquid and vapour states, an enhanced sampling method with grand canonical biasing would be required.

Fortuitously, water in the open state structure (PDB ID: 6DG8) exhibits a more interesting behaviour amenable to a quantitative comparison of water models. Here, both the Drude oscillator model and the additive simulation predict significant de-wetting in the hydrophobic region of the pore (Figure 5.5A). However, the average water density in the hydrophobic gate predicted by the mTIP3P and SWM4-NDP models is approximately half the density of bulk water and thus significantly higher than in the closed and intermediate states of the receptor, where the water density is close to zero. As shown in Figure 5.5B, both these models undergo stochastic fluctuations between the liquid and vapour states of the pore.

By contrast, in simulations of the open state structure employing the AMOEBA polarisable force field, water strongly favours the liquid state. The time-averaged water

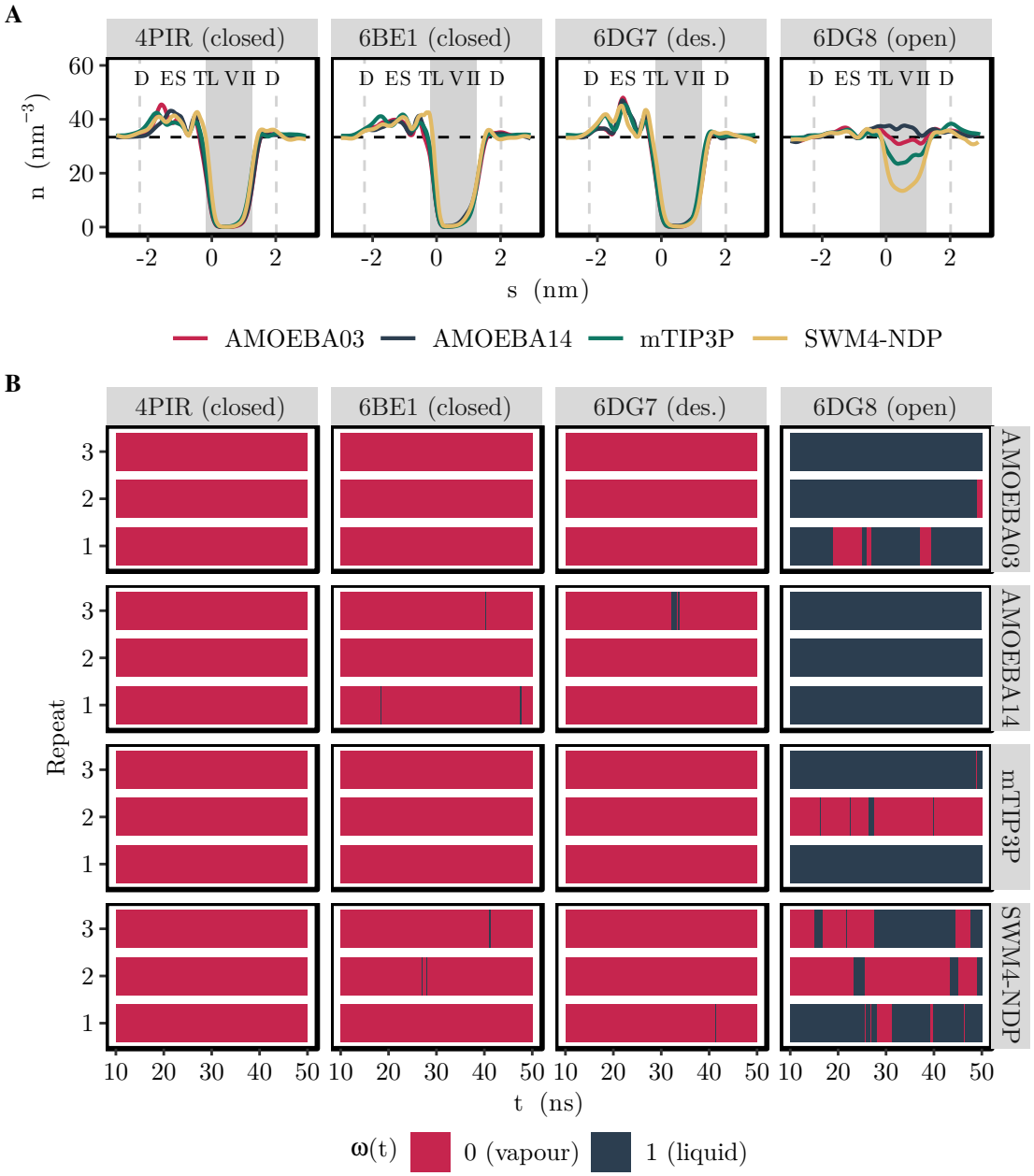



**Figure 5.5:** Comparison of pore hydration using polarisable force fields and water models. The additive mTIP3P water model is also shown as a reference case. (A) Water density in four conformational states of the 5-HT$_3$ receptor. The profiles shown represent the average of three independent simulations (five in the case of mTIP3P) and the dashed horizontal line indicates the density of bulk water. Dashed vertical lines and the shaded background denote the extent of the protein and the hydrophobic gate region respectively. Single letter codes for the pore-facing amino acid sidechains are given at the top of each panel. (B) Beckstein openness over time for three independent repeats. Repeats 4 and 5 are not shown for the mTIP3P water model and the final 100 ns of the mTIP3P and SWM4-NDP simulations are omitted for clarity.

density is close to that of bulk water throughout the entire channel pore (Figure 5.5A) and water exists in the vapour state for a significant amount of time only during one simulation with the AMOEBA03 model (Figure 5.5B). In all three simulations with the AMOEBA14 water model, the hydrophobic gate does not de-wet and water exists exclusively in the liquid state. Such a fully hydrated pore would be expected for the open state of the channel and it is intriguing that this prediction is made by what is arguably the most accurate water model available (Figure 5.1).

It is important to note that due to their high computational cost, the AMOEBA simulations were carried out for only 50 ns and it cannot be ruled out that significant de-wetting of the hydrophobic gate would occur in a longer simulation. In principle, it is possible that the explicit inclusion of atomic polarisation slows down the dynamics of de-wetting, thereby creating a metastable liquid state in a system that actually prefers the vapour state. However, the observation that one of the AMOEBA03 repeats exhibits multiple transitions between the liquid and vapour states within 50 ns suggests that this is not the case, lending credibility to the conclusion that the AMOEBA14 model predicts a completely hydrated hydrophobic gate. This is further supported by the fact that fluctuations of Beckstein openness in simulations employing the similarly polarisable Drude oscillator force field also occur on a time scale of $\ll$50 ns (Figure 5.5B).

It is now worthwhile to investigate whether the alternative additive water models exhibit a qualitatively similar behaviour. Figure 5.6 examines the pore hydration behaviour in simulations employing the OPC, TIP4P, TIP4P/2005, and SPC/E water models in conjunction with the CHARMM36m force field. As can be seen, simulation systems based on the closed (PDB ID: 4PIR and 6BE1) and intermediate state structures (PDB ID: 6DG7) mirror the behaviour observed in polarisable MD simulations.

Water density is near zero in the hydrophobic gate region and close to the bulk value throughout the rest of the pore, irrespective of the water model (Figure 5.6A). In the hydrophobic section of the pore, water exists in a vapour state for the entire duration of

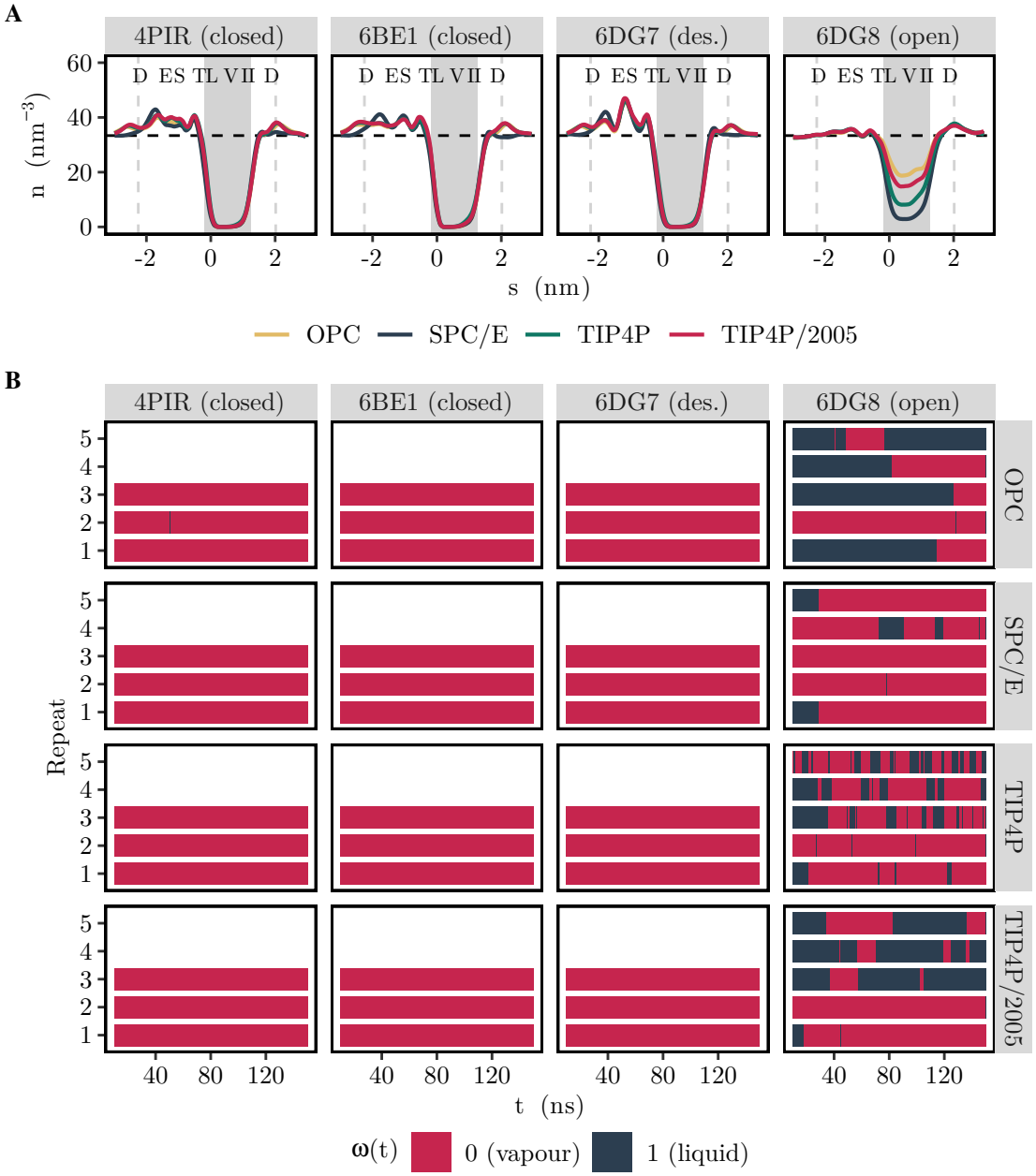



**Figure 5.6:** Comparison of pore hydration using alternative additive water models. (A) Water density in four conformational states of the 5-HT$_3$ receptor. The profiles shown represent the average of three independent simulations (five in the case of the open state (PDB ID: 6DG8) structure) and the dashed horizontal line indicates the density of bulk water. A shaded background represents the hydrophobic gate region and dashed vertical lines denote the extent of the protein. Single letter codes for the pore-facing amino acid side chains are shown at the top of each panel. (B) Beckstein openness over time shown individually for each repeat.

nearly all 150 ns long simulations and very short wetting events occur only once during the second repeat of the OPC and TIP4P simulations. This further demonstrates that a hydrated pore is energetically strongly disfavoured, if the channel assumes the closed or intermediate state conformations.

In contrast to this, liquid-vapour transitions can be observed with all additive water models in simulations of the open state 5-$HT_3$ receptor structure (PDB ID: 6DG8). For this conformation of the channel, the time-averaged water density in the hydrophobic gate region takes intermediate values between zero and the density of bulk water. For the OPC and TIP4P/2005 models, the density profile exhibits a minimum slightly higher than half the bulk density, while for the TIP4P and SPC/E models the minimum density falls below this value (Figure 5.6A).

The openness time series displayed in Figure 5.6B correspondingly illustrates that the OPC and TIP4P/2005 water models spend more time in the liquid state than the TIP4P model, while the SPC/E model exists primarily in the vapour state. Notably, the second simulation of each water model often seems to favour the vapour state more strongly than other repeats using the same model. This can also be observed for the mTIP3P water model (Figure 5.5B), suggesting a systematic effect, as the second repeat of all water models originates from the same coarse-grained self-assembly and atomistic equilibration simulations (see Section 5.1 for the details of the setup pipeline).

Interestingly, the time scale on which fluctuations occur also depends on the water model. For example, TIP4P water undergoes frequent phase transitions, with both the liquid and vapour states typically lasting for less than ∼10 ns. By contrast, in simulations with the OPC model the liquid state can survive for more than 100 ns before the first transition to the vapour state occurs. Slow transitional dynamics can also be observed for the TIP4P/2005 water model, where in some repeats the liquid state persists for up to ∼40 ns. The SPC/E model on the other hand favours the vapour state and only a small number of phase transitions occur. When the system does enter the liquid state,

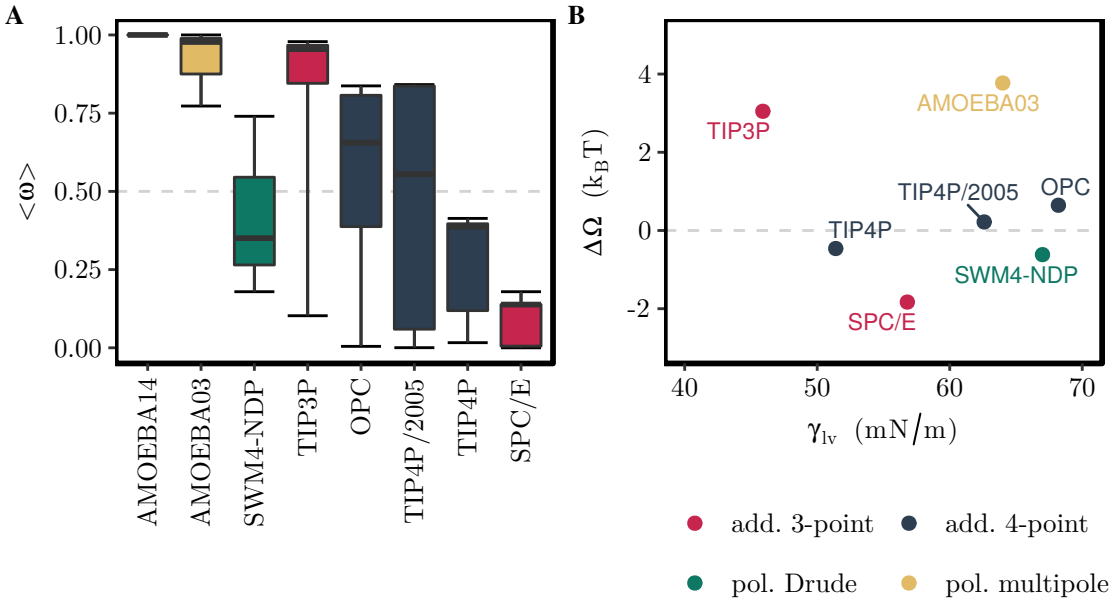



**Figure 5.7:** Hydration equilibrium of the open state (PDB ID: 6DG8) 5-$HT_3$ receptor pore predicted by different water models. (A) Time-averaged Beckstein openness, $\langle\omega\rangle$, is shown as a measure for the fraction of simulation time during which the channel pore is hydrated. The boxplot shows the median and quartiles across five independent repeats for the additive and three independent repeats for the polarisable water models. (B) Dependence of the free energy difference, $\Delta\Omega$, between the liquid and vapour states of the channel pore on the liquid-vapour surface tension of the water model, estimated from the median openness over all repeats. For the AMOEBA14 model, no free energy could be calculated as the pore never enters the vapour state. The dashed grey line represents the case of equal probability for the liquid and vapour states.

it does not survive for more than ∼20 ns.

It is unclear, however, whether these different survival times indicate inherently different time scales for the de-wetting and re-wetting processes or if they are a consequence of the different equilibrium constants for the liquid-vapour transition. This question will be addressed in Chapter 6, where the application of an external electric field is used to drive the system to a state point with equal probability for the liquid and vapour phases.

With the exception of AMOEBA14, all water models undergo liquid-vapour transitions in the pore of the open conformation of the 5-$HT_3$ receptor. Simulations based on the 6DG8 structure can therefore be used to establish a ranking of water models with respect

to their relative proclivity for the liquid or vapour states. Figure 5.7A displays the time-averaged openness for each water model. As can be seen, the two polarisable AMOEBA models strongly favour the liquid state, while the polarisable Drude oscillator model SWM4-NDP assigns roughly equal probability to both the liquid and vapour states.

Amongst the additive models, mTIP3P is the most likely to predict a hydrated pore, while with the TIP4P and SPC/E models the hydrophobic gate exists predominantly in the vapour state. Intermediate openness values are obtained for the OPC and TIP4P/2005 models. It should be noted that the OPC, mTIP3P, TIP4P, and TIP4P/2005 models exhibit a skewed distribution featuring a long tail towards lower values of $\langle\omega\rangle$. This is a consequence of the tendency of the second repeat to de-wet more strongly than the other repeats of these water models, but does not strongly impact the median openness, on which this ranking is based.

From the time-averaged openness, the free energy difference between the liquid and vapour states of the pore can be estimated. Before doing so, it is expedient to recall the thermodynamic Beckstein model[44] for the energetics of hydrating a cylindrical nanopore (Section 1.2.4). Since the pore radius of the 5-$HT_3$ receptor does not vary significantly throughout the transmembrane domain, Equation 1.4 is applicable here. Crucially, the Beckstein model predicts a free energy difference proportional to the liquid-vapour surface tension of the solvent and this proportionality would be preserved even in a model that more accurately captures the geometric details of the 5-$HT_3$ receptor pore.

Figure 5.7B therefore relates the energetic cost of pore hydration to the liquid-vapour surface tension of the respective water model. It is clear that there is no proportionality between these two quantities, which in fact appear to be largely uncorrelated. As all simulations were carried out in the same channel pore, geometric differences cannot explain this discrepancy. Consequently, the deviation from the expected proportionality must be attributed to the contact angle, $\cos\theta_c$, i.e. to the effective hydrophobicity of the pore wall.

It is important to note that the contact angle is not a property of the water model itself and instead arises from the interplay between water model and protein force field. Since all common protein force fields are parameterised against an existing water model (their host water model), the value of $\cos\theta_c$ is primarily determined by the protein parameterisation. To decide whether a water model is suitable to study hydrophobic gating, it is therefore not sufficient to only assess the accuracy of its liquid-vapour surface tension. This is well illustrated by the fact that the mTIP3P and AMOEBA14 water models both predict a large Beckstein openness, despite having the largest difference in liquid-vapour surface tension of all water models considered here (Figure 5.1).

A direct consequence of this is that modern additive water models like OPC and TIP4P/2005 are unlikely to correctly predict the hydration state of a channel pore, despite their significantly improved accuracy in describing the properties of pure water. Combining these models with protein force fields parameterised against older water models may result in a wrong value for the effective hydrophobicity of the pore. While protein force field parameterisation does not typically target the contact angle directly, it is implicitly considered in the calculation of hydration free energies for amino acid analogues. For this reason, the mTIP3P water model in conjunction with an appropriate force field like CHARMM36m can be expected to yield more accurate results, as insufficiencies of the water model can be cancelled out by the protein parameter set.

Of the models shown in Figure 5.7, only AMOEBA03, SWM4-NDP, and mTIP3P were used together with a directly corresponding protein force field. Interestingly, the AMOEBA03 and mTIP3P simulations agree well with one another in predicting a pore that is mostly hydrated and the estimated energetic cost of pore hydration differs by less than $1\,k_BT$ between these models. This prediction is also credible from a structural point of view, as the 6DG8 structure captures the 5-$HT_3$ receptor in a serotonin-bound conformation, for which a hydrated hydrophobic gate would be expected. However, predictions made by the Drude oscillator force field deviate from the AMOEBA and

mTIP3P results by $\sim$4 $k_BT$ and it is unclear what causes this deviation.

Finally, it needs to be pointed out that the strong influence of the contact angle also implies a strong dependence on the chemical character of the pore surface. In a channel with a pore lining composed of amino acid residues other than the leucine and valine residues lining the hydrophobic gate of the 5-$HT_3$ receptor, the ranking of water models could well be different from what is shown in Figure 5.7. Further research is needed to elucidate if this is indeed the case.

### Polarisation Effects

Having established that the free energy of pore hydration can vary by as much as $\sim 5\, k_BT$ across different water models, this section aims to address the question of how relevant polarisation effects are for an accurate description of water confined in the hydrophobic gate of an ion channel. Figure 5.8 shows how the magnitude and orientation of the molecular dipole moment of water vary along the pore of the 5-$HT_3$ receptor. As in the case of pore hydration discussed in the previous section, the closed (PDB ID: 4PIR and 6BE1) and intermediate (PDB ID: 6DG7) state structures yield very similar results.

In these three structures, both AMOEBA models exhibit a sharply decreasing dipole moment in the hydrophobic gate region of the pore (Figure 5.8A). Here the dipole moment drops from $\sim$2.7 D in the bulk to a minimum of $\sim$2.2 D between the 9′ leucine and 13′ valine residues. This represents a remarkable change, as the dipole moment is effectively reduced to a value close to the dipole moment of gas phase water, while elsewhere in the pore the dipole moment does not deviate markedly from the bulk value. Both models also agree well with one another, except for some deviations in the hydrophobic gate region, where sampling is relatively poor due to the low density of water molecules in this region.

Notably, the effect of the hydrophobic gate is much larger than that of charged residues lining the pore. For example, in the closed state structure (PDB ID: 4PIR), the dipole

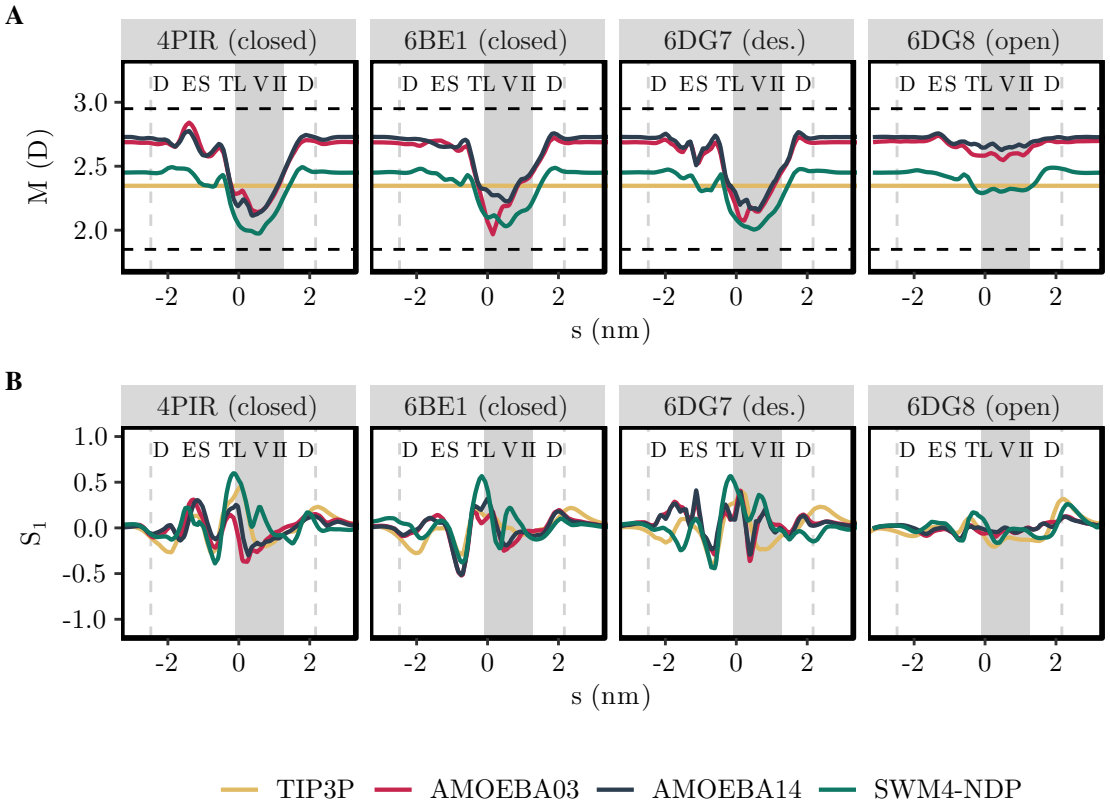



**Figure 5.8:** Molecular dipole moment of water in the pore of four conformational states of the 5-$HT_3$ receptor. (A) Magnitude of the molecular dipole moment, $M$. (B) Orientational order parameter of molecular dipole moment, $S_1$. The shaded area represents the hydrophobic gate region, in which de-wetting occurs. Dashed vertical lines indicate the extent of the simulated protein structure. Dashed horizontal lines at 1.85 D and 2.95 D represent the dipole moment of water in the gas and liquid phase respectively. Single letter codes for the pore-facing amino acid side chains are given at the top of each panel. Curves shown represent the average over three independent repeats.

moment of both AMOEBA models peaks near the $-1'$ glutamate residue (Figure 5.8A). However, the magnitude of this peak is only a fraction of the decrease experienced in the hydrophobic gate. In the other two de-wetted structures (PDB ID: 6BE1 and 6DG7), this peak does not exist at all. In part, this can be explained by the fact that the charges on the five glutamate side chains are arranged in a ring around the permeation pathway. Consequently, their net field will cancel in the centre of the pore and a water molecule placed here will experience negligible lateral electrostatic forces. Nonetheless,

this emphasises the importance of the hydrophobic nature of the 9′ leucine residues as the main gate of the 5-$HT_3$ receptor.

In the open state structure (PDB ID: 6DG8), on the other hand, the decrease in dipole moment throughout the hydrophobic section of the pore is relatively minor and both AMOEBA models exhibit only small fluctuations around the bulk value (Figure 5.8A). This corresponds well to the previous observation that in the serotonin-bound conformation of the channel, the hydrophobic gate remains fully hydrated during simulations with the AMOEBA force field. Even the intermittent de-wetting events occurring during the third AMOEBA03 repeat (Figure 5.5B) are reflected here by a slightly lower dipole moment.

The nearly constant dipole moment profile in the open state structure also suggests that the sharp decrease in dipole moment in the closed and intermediate state structures is not directly caused by the presence of hydrophobic amino acids, but is instead due to the absence of neighbouring water molecules. The hydrophobicity of the pore surface is very similar in all structures of the 5-$HT_3$ receptor, but in the open state structure the dipole moment of water decreases by no more than ∼0.1 D even in direct proximity to a ring of leucine residues.

This implies that water can sustain a near bulk-like dipole moment in the vicinity of a hydrophobic protein surface and only assumes a gas-phase-like dipole moment when it exists in isolation from other water molecules. The fact that in the wider open state pore water molecules will on average have a larger distance from the pore wall does not change this conclusion, as it cannot explain the five-fold larger reduction in dipole moment observed in the de-wetted pores of the closed and intermediate channel conformations.

Results obtained from simulations employing the Drude oscillator force field present a very similar picture. While in the bulk regime, SWM4-NDP water has a dipole moment of ∼2.5 D, this decreases to ∼2.0 D in the hydrophobic gate of the closed and intermediate conformations of the 5-$HT_3$ receptor (Figure 5.8A). In the open conformation, the Drude

model again mirrors the AMOEBA models and exhibits only a minor decrease in dipole moment throughout the hydrophobic section of the pore. In general, the dipole profile of the SWM4-NDP model behaves like the AMOEBA profiles shifted down by $\sim$0.25 D, reflecting the lower bulk water dipole moment of the Drude model.

Figure 5.8B shows how the orientation of the dipole moment changes as a water molecule passes through the 5-$HT_3$ receptor. In the open state structure (PDB ID: 4PIR), there is no strong orientational preference anywhere in the pore, while in the three de-wetted structures (PDB IDs: 4PIR, 6BE1, and 6DG7) orientational ordering is apparent. The orientational order parameter peaks near the $-1'$ glutamate and $6'$ threonine residues and falls below its bulk value in the vicinity of the $2'$ serine residue. Otherwise, the order parameter is close to zero, indicating that the dipole moment either has no preferential orientation or is oriented perpendicular to the transmembrane direction. More generally, the orientational behaviour of the dipole moment is consistent across all three polarisable models and is similarly captured by the additive mTIP3P model.

It is also worthwhile to consider which components of the molecular dipole moment contribute to the observed variation inside the hydrophobic gate. As can be seen from Figure 5.9, the only component that varies significantly along the pore is the induced atomic dipole moment. This means that while AMOEBA water is flexible, there is no systematic variation of its average bond length or the H–O–H angle as a water molecule moves through the channel pore. The primary difference between the AMOEBA03 and AMOEBA14 models is the relative strength of the partial charges and permanent atomic dipoles and for both models the induced dipole moment varies by $\sim$50 %. Since the SWM4-NDP model is rigid, only the induced dipole moment can vary in this case. However, it is worth pointing out that this component nearly decreases to zero inside the hydrophobic gate (Figure 5.9), i.e. the Drude dipole moment is reduced by almost the maximum amount possible.

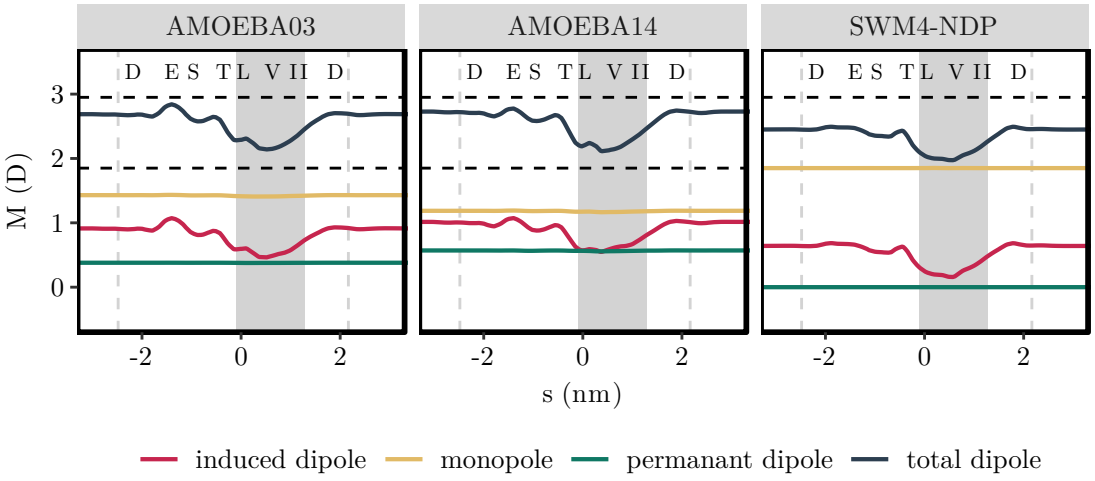



**Figure 5.9:** Individual components of the molecular dipole moment of water in the pore of the closed state (PDB ID: 4PIR) structure of the 5-$HT_3$ receptor. Profiles shown represent the average of three independent simulations for each of the polarisable AMOEBA03, AMOEBA14, and SWM4-NDP water models. The shaded area indicates the hydrophobic gate region and the dashed vertical lines denote the region where water is confined within the protein pore. Dashed horizontal lines represent the dipole moment of water in the liquid and gas phases of 2.95 D and 1.85 D respectively. Single letter codes indicating the location of the pore-facing amino acids are shown at the top of each panel.

Given that the theoretical models of polarisation underlying the CHARMM Drude and AMOEBA force fields are very different, the great similarity of the dipole moment profiles is remarkable. While not in exact quantitative agreement, both approaches predict a significant variation for the magnitude of the dipole moment of water in the hydrophobic gate region of the 5-$HT_3$ receptor, giving a strong indication that this reflects the actual physical behaviour of water dipoles in hydrophobic confinement and that it is not an artefact of a particular polarisation model or force field parameterisation.

It is also important to note that this effect cannot be captured by an additive water model. As illustrated in Figure 5.8A, mTIP3P water has a dipole moment of ∼2.4 D, roughly corresponding to the average of liquid and gas phase water. This does not accurately reflect the prediction of the polarisable models, whose dipole moment (in the closed and intermediate channel conformations) lies below this value in the hydrophobic

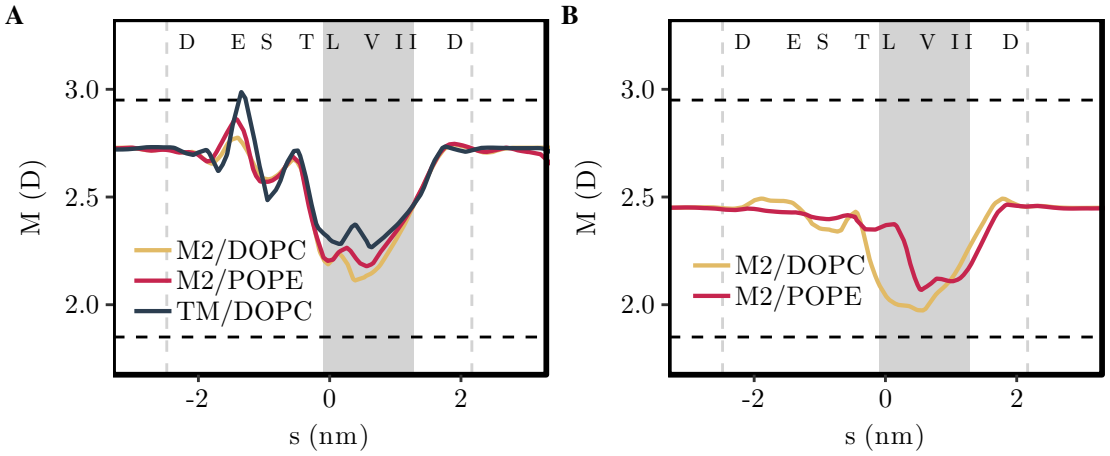



**Figure 5.10:** Magnitude of the molecular dipole moment of water in three different simulation systems based on the closed state (PDB ID: 4PIR) structure of the 5-$HT_3$ receptor. Simulations employed (A) the AMOEBA13 protein force field together with the AMOEBA14 water model and (B) the CHARMM Drude force field combined with the SWM4-NDP model for water. Curves shown represent the average over three independent repeats. For both force fields, the receptor was represented by either the M2 helix bundle embedded in a pure DOPC or pure POPE lipid bilayer or by the complete transmembrane domain embedded in a DOPC bilayer.

gate and exceeds it elsewhere. Other additive models would also predict a constant dipole moment profile, albeit at slightly different absolute values.

One question that remains open is whether the dipole moment is sensitive to the overall electrostatic environment of the protein. For computational efficiency, the 5-$HT_3$ receptor was so far represented only by its pore-lining M2 helices, whereas in reality the transmembrane domain comprises additional rings of $\alpha$-helices that shield the pore from the influence of the lipid bilayer. In principle it is possible that this simplified structural model creates a nonphysiological electrostatic environment inside the channel pore that causes artefacts in the dipole moment profile.

To investigate this possibility, two additional sets of simulations based on the closed state 5-$HT_3$ receptor structure (PDB ID: 4PIR) were performed. In the first one, the protein was still represented by the M2 helix bundle, but was embedded in a POPE rather than a DOPC bilayer. In the second, the complete transmembrane (TM) domain, embedded in a DOPC bilayer, was simulated. Due to technical limitations of the

DrudePrepper tool[†], simulations of the complete TM domain could not be carried out with the CHARMM Drude force field.

Throughout these additional simulations the hydrophobic gate remained de-wetted in agreement with what had been observed previously for the closed channel conformation. Figure 5.10 compares the corresponding dipole moment profiles to the baseline simulations. As can be seen, all three profiles derived from the AMOEBA force field are in near-quantitative agreement, featuring a major decrease in the hydrophobic section of the pore and two minor peaks near the $-1'$ glutamate and $6'$ threonine residue rings. The Drude simulation with POPE lipids predicts a higher dipole moment near the $9'$ leucine residues, but still exhibits a significant decrease in the remainder of the hydrophobic gate. This indicates that the dipole moment of water inside the pore is primarily influenced by short-ranged interactions and confirms that it does not critically depend on the complete representation of the transmembrane domain.

Finally, Figure 5.11 summarises the behaviour of the water dipole in the pore of the 5-$HT_3$ receptor as predicted by both the AMOEBA and CHARMM Drude models. In the bulk domain either side of the channel, water assumes a large liquid phase dipole moment with no orientational preference. Entering the pore from the intracellular side, the dipole moment is enhanced near the $1'$ ring of glutamate residues, aligns itself with the transmembrane direction and points towards the extracellular domain. Further up in the channel its magnitude returns to the bulk value and its orientation is flipped. As water approaches the $9'$ ring of leucine residues, the magnitude of the dipole moment begins to decrease markedly and its orientation reverses again. In the middle of the hydrophobic gate region the dipole moment approaches a value similar to that of gas phase water and loses its orientational preference.

[†]The definition of the PDB file format reserves five characters for each atom index in `CONECT` records, which are required for the specification of bonds in non-standard residues like lipid molecules. In the TM domain simulation system, the number of atoms exceeds 99 999 so that bonds cannot be uniquely defined in a PDB file. This then causes DrudePrepper to fail.

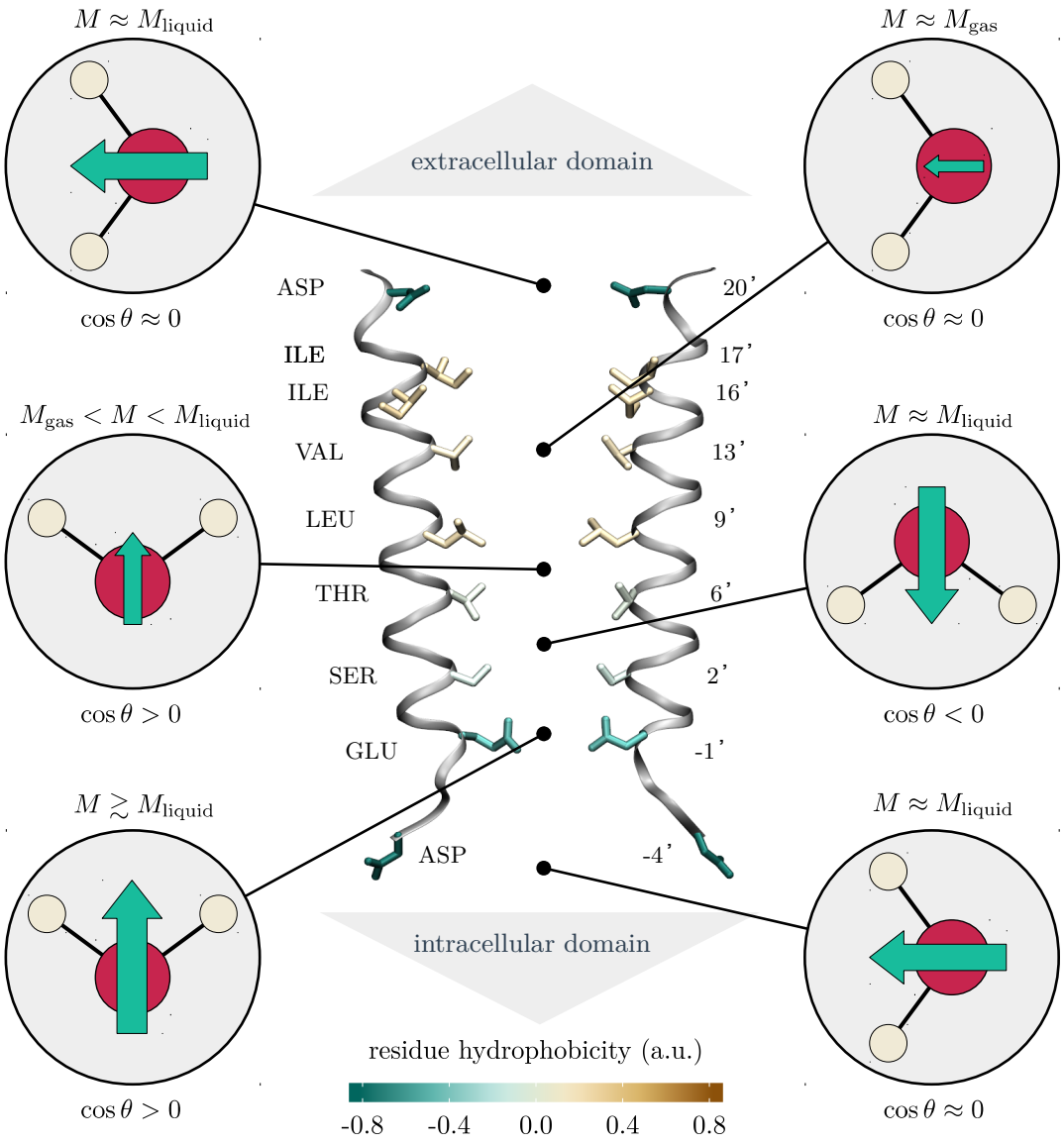



**Figure 5.11:** Schematic summary of water dipole moment behaviour in the closed state (PDB ID: 4PIR) structure of the 5-$HT_3$ receptor. Two of the M2 helices are shown as ribbons with pore-facing residues displayed in licorice representation and coloured by hydrophobicity. At six positions along the channel pore, the characteristic magnitude and orientation of the molecular dipole moment of water are indicated by a green arrow representing the net dipole vector. In the hydrophobic pore section between the 9′ leucine and 16′ isoleucine residue rings, the dipole moment is significantly reduced compared to its value in bulk water. Here the water dipole resembles that of gas phase water molecules.

### 5.2.3 Influence of Polarisation on Ion Conduction

In the previous section, it was shown that the molecular dipole moment of water can be reduced significantly in the hydrophobic gate of an ion channel, giving a strong indication that induced polarisation needs to be modelled explicitly to obtain a quantitatively accurate picture of biomolecular transport processes. Furthermore, it was shown that the choice of water model has a significant influence on the hydration state of the channel pore. This section therefore explores how the behaviour of ions is impacted by induced polarisation and what role the solvent model plays in ion conduction.

To this end, umbrella sampling simulations were carried out to obtain a potential of mean force (PMF) profile for ion permeation in the open state 5-$HT_3$ receptor (PDB ID: 6DG8). These simulations employ the polarisable AMOEBA and CHARMM Drude force fields with the AMOEBA14 and SWM4-NDP water models respectively, as well as the additive CHARMM36m force field with the mTIP3P and TIP4P/2005 water models. PMF profiles were calculated for both $Na^+$ and $Cl^-$ ions as examples of biologically relevant cation and anion species. In addition, radial distribution functions were used to quantify the structure of the hydration shell around these ions as they pass through the channel pore.

#### Potential of Mean Force Calculated with the CHARMM Additive Force Field

Figure 5.12 compares the free energy profiles across ion species and water models, while Figure 5.13 displays diagnostic plots for assessing umbrella histogram overlap and PMF convergence. As can be seen in Figure 5.12, the additive mTIP3P and TIP4P/2005 models yield qualitatively similar free energy profiles for both $Na^+$ and $Cl^-$ ions. For these additive force fields, the umbrella histogram exhibits good overlap (Figure 5.13A) and the error metrics fall below the threshold of $1\,k_BT$ well within the 5 ns equilibration period, indicating that the resulting PMF profiles are converged (Figure 5.13B–C).

In the case of $Na^+$, the PMF is characterised by a well near the 20′ ring of negatively

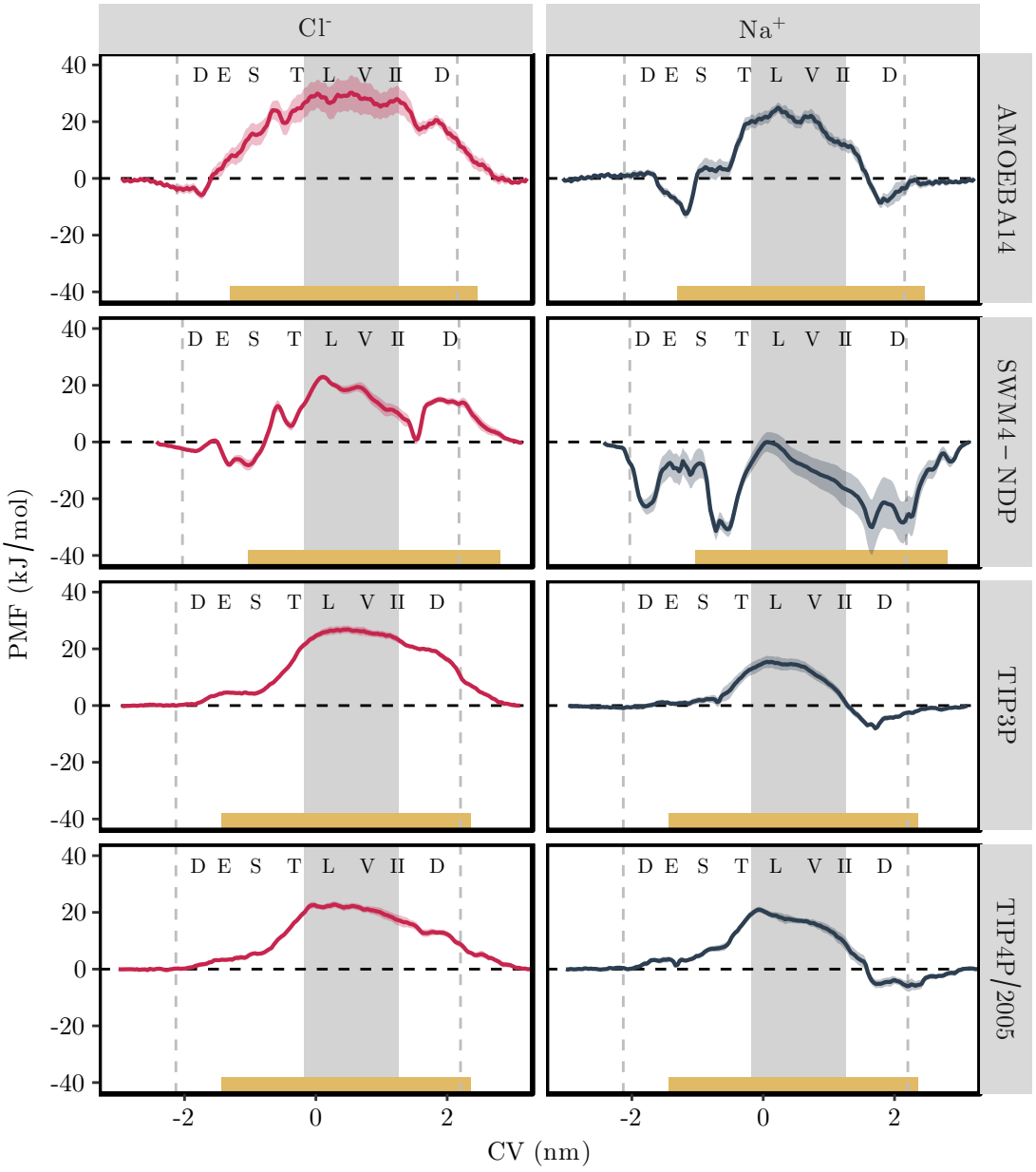



**Figure 5.12:** Single ion PMF profiles for $Na^+$ or $Cl^-$ ions in the pore of the open state 5-$HT_3$ receptor (PDB ID: 6DG8). The solid line indicates the free energy profile calculated from the final 5 ns of each umbrella window. Confidence bands were obtained by calculating the standard error over independent 1 ns long sampling blocks during the same time period. Vertical dashed lines denote the extent of the protein and the yellow bar represents the position of the lipid bilayer. The hydrophobic gate region, where de-wetting occurs in equilibrium simulations, is denoted by the grey background shading. Single letter codes for the pore-facing amino acid side chains are given at the top of each panel. Dashed horizontal lines indicate the bulk energy. Note that the $Na^+$ PMF calculated with the SWM4-NDP water model is not converged (see main text for details).

charged aspartic acid residues and an extensive free energy barrier in the hydrophobic gate region. Interestingly, there is no energy well in the vicinity of the $-1'$ glutamic acid residues. This can potentially be explained by the entropic cost incurred due to the stricter confinement near the glutamic acid, where the pore radius is $\sim$0.24 nm, compared to 0.60 nm at the site of the $20'$ aspartic acid. For both water models, the energy barrier extends from the $6'$ threonine to the $16'$ isoleucine residues, but with TIP4P/2005 water there is a distinct peak between the $6'$ threonine and $9'$ leucine residues. Quantitatively, this corresponds to an energy barrier of $(21.0 \pm 1.1)\,\mathrm{kJ\,mol^{-1}}$ for the TIP4P/2005 model, which is significantly higher than the value of $(15.6 \pm 2.2)\,\mathrm{kJ\,mol^{-1}}$ predicted by the mTIP3P model.

$Cl^-$ ions also experience the largest energy barrier in the hydrophobic gate region, with a peak height of $(26.9 \pm 1.4)\,\mathrm{kJ\,mol^{-1}}$ when the mTIP3P model is used and a peak height of $(22.8 \pm 1.1)\,\mathrm{kJ\,mol^{-1}}$ if the TIP4P/2005 model is used. Both water models thus predict a larger barrier for $Cl^-$ ions than for $Na^+$ ions, but the difference in barrier heights is much larger for the mTIP3P model. Physiologically, this is sensible, as the 5-$HT_3$ receptor is known to be cation selective, although the selectivity of the channel could in principle be caused by residues outside the transmembrane domain as well.[225] The presence of $Cl^-$ is also energetically disfavoured near the $20'$ ring of aspartic acid residues, but the energetic cost due to the negative charges in this region is lower than that caused by the hydrophobic gate. As in the case of $Na^+$, the $-1'$ ring of glutamic acid residues appears to have a much smaller effect, and the free energy profile of $Cl^-$ ions is only $\sim$5 $\mathrm{kJ\,mol^{-1}}$ above bulk level in this region.

#### Potential of Mean Force Calculated with the AMOEBA Force Field

Free energy profiles derived from simulations with the AMOEBA polarisable force field exhibit notable qualitative differences from the additive cases (Figure 5.12). For $Na^+$ ions, the PMF profile still peaks in the hydrophobic gate region with a maximum of

$(25.0 \pm 1.9)\,\mathrm{kJ\,mol^{-1}}$ near the $9'$ leucine residues. However, it exhibits two energy wells, one near the $20'$ aspartic acid residues and the second at the position of the $-1'$ glutamic acid residues. In addition, the free energy profile appears less smooth than the PMFs obtained using additive force fields and exhibits more secondary features on a smaller spatial scale. For example, there are two distinct peaks near the sites of the $9'$ leucine and $13'$ valine residues. This suggests that for a polarisable ion there are more specifically favourable or disfavourable positions along the channel pore, depending on whether it can create induced dipoles in neighbouring atoms.

On the whole, the $Cl^-$ PMF derived from the AMOEBA force field resembles the profile predicted by the mTIP3P water model. It peaks in the hydrophobic gate region at a value of $(30.20 \pm 6.20)\,\mathrm{kJ\,mol^{-1}}$ and also disfavours a placement of $Cl^-$ near the $20'$ aspartic acid residues. In contrast to the additive case, however, the AMOEBA force field predicts a more gradual increase of the PMF and a larger free energy near the $-1'$ glutamic acid residues. It also predicts a small energy well in the vicinity of the $-4'$ ring of aspartic acid residues. While counterintuitive, this is likely due to the fact that the charged sidechain of the $-4'$ residues does not point directly towards the pore (Figure 5.11), so that the $Cl^-$ ion can favourably interact with the backbone atoms, in which it may induce additional polarisation. As in the case of $Na^+$, the PMF appears to be less smooth than its additive counterparts, although some of its secondary features are smaller than the estimated statistical uncertainty of the profile.

As can be seen in Figure 5.13A, umbrella sampling with the AMOEBA force field results in a histogram that is less even than in the additive cases, but still produces satisfactory overlap between windows. However, convergence of the PMF appears to require more time than for an additive force field (Figure 5.13A–B). For both $Na^+$ and $Cl^-$, the deviation of the instantaneous PMF from the final free energy profile starts out larger and initially even increases, before ultimately falling below the threshold of $1\,k_BT$. For $Na^+$, this is achieved within $\sim$5 ns, while for $Cl^-$ full convergence may only

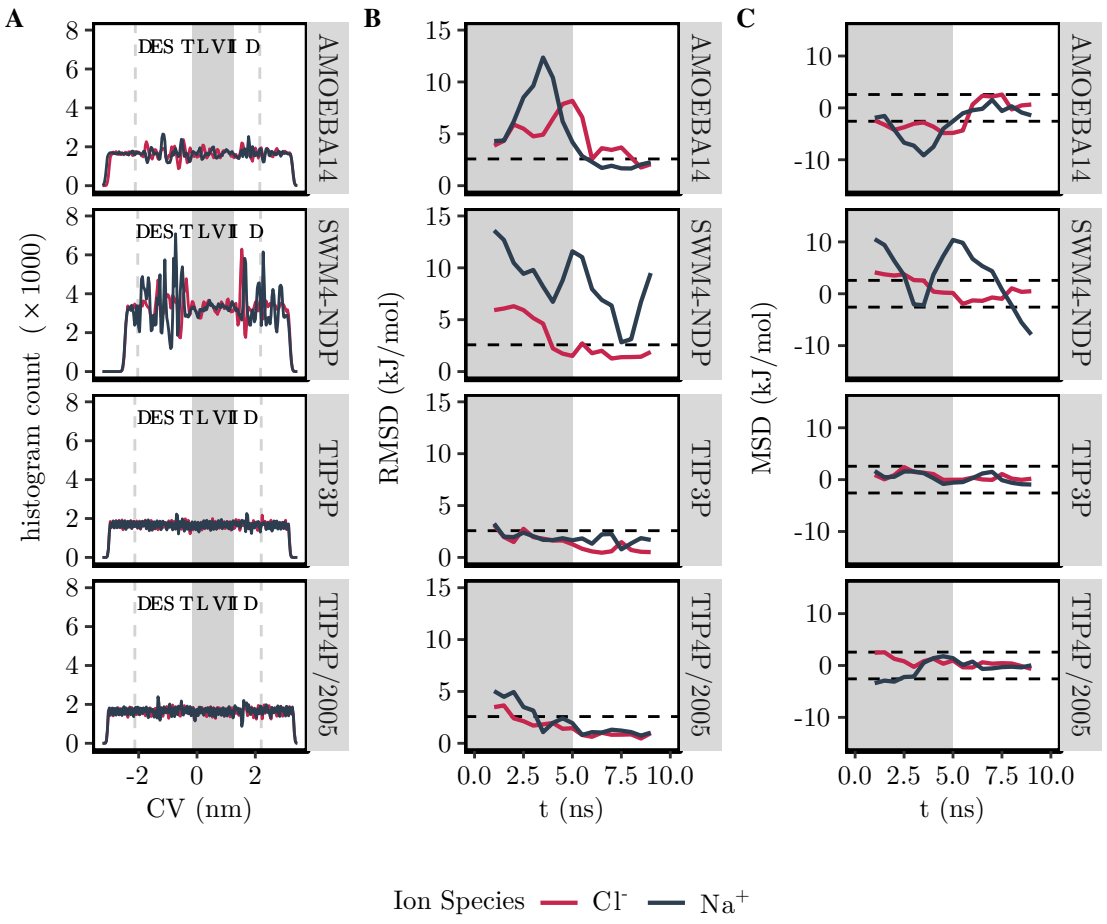



**Figure 5.13:** Convergence of umbrella sampling simulations. (A) Histogram for both $Na^+$ and $Cl^-$ ions over all umbrella sampling windows. Dashed grey lines indicate the extent of the protein and the shaded area denotes the hydrophobic gate region that de-wets in equilibrium simulations. Single letter codes for the pore-facing amino acids are shown at the top of each panel. Due to a finer spacing of umbrella windows, the absolute number of samples is larger for the Drude oscillator model SWM4-NDP. (B,C) Convergence of potential of mean force evaluated in terms of the root mean squared deviation (B) and mean signed deviation (C) with respect to the final profile. The shaded background indicates the 5 ns long equilibration period which is discarded in the calculation of the final PMF shown in Figure 5.12. Dashed horizontal lines indicate where the deviation falls below the thermal energy of $1\,k_BT$.

be reached after ∼6 ns.

The slower equilibration in the polarisable case could be an indication that the restrained ion explores its lateral degrees of freedom more slowly. The ionic charge may create induced dipoles on the pore atoms, which would in turn cause the ion to bind

more tightly to otherwise nonpolar residues. As a consequence, the ion would tend to closely associate with individual residues and move between different subunits of the channel protein less frequently. Beyond the need for a longer equilibration period in umbrella sampling, this effect may also be relevant for the dynamics of ion conduction. A polarisable ion that interacted with the pore walls more strongly would experience a larger effective friction and would consequently be transported through the channel at a lower rate than a non-polarisable ion, provided that the driving force was identical.

### Potential of Mean Force Calculated with the CHARMM Drude Force Field

Free energy profiles derived from the CHARMM Drude force field also differ markedly from the additive simulations (Figure 5.12). The $Na^+$ PMF in particular stands out by predicting a free energy below the bulk water value throughout the entire pore. This is implausible as the lateral confinement of the ion inside the pore should always incur an entropic cost and both the additive CHARMM36m and polarisable AMOEBA simulations suggest a significant energy barrier. As shown in Figure 5.13, the $Na^+$ PMF derived from the CHARMM Drude simulations is not converged even after 10 ns. Both error metrics repeatedly decrease and increase and the MSD (Figure 5.13C) even changes sign.

This lack of convergence could be related to an insufficient histogram overlap. As can be seen in Figure 5.13A, the histograms for the umbrella sampling simulations with the SWM4-NDP water model are highly variable, although each bin nevertheless contains at least 1000 samples. This behaviour is remarkable, because for the CHARMM Drude force field, the window distance has already been reduced to 0.05 nm from a value of 0.10 nm used with both the AMOEBA and CHARMM36m force fields. Further potential problems, such as a dynamical instability of the simulation system or the intrusion of lipid tails into the pore, have been ruled out by visual inspection of all umbrella windows. It thus remains unclear why the CHARMM Drude force field fails to converge to a

physically sensible PMF for $Na^+$ ions. Future studies may therefore explore the utility of alternative free energy techniques, which do not rely on predefined sampling windows, such as replica exchange or metadynamics.

As can be seen in Figure 5.12, for $Cl^-$ ions the CHARMM Drude force field predicts a primary barrier in the hydrophobic gate region with a peak height of $(22.9 \pm 0.8)\,\mathrm{kJ\,mol^{-1}}$, similar to the predictions made by the other force fields. A further similarity is the high free energy cost for placing an ion near the $20'$ ring of aspartic acid residues. However, in contrast to all other force fields, the Drude model predicts an energy well near the $-1'$ glutamic acid residues. This well is both distinct from and deeper than the negative free energy observed near the $-4'$ aspartic acid residues in simulations with the AMOEBA force field.

As an additional noteworthy feature, the $Cl^-$ PMF exhibits two local minima, near the $6'$ threonine residues and between the $17'$ isoleucine and $20'$ aspartic acid residues respectively. Similar energy wells are also present in the corresponding free energy profile calculated with the AMOEBA force field, but are much more pronounced when the CHARMM Drude force field is used. Taken together, these observations support the hypothesis that the inclusion of polarisation effects creates a more rugged free energy landscape than what is predicted by additive force fields.

By comparison to $Na^+$, the $Cl^-$ PMF derived from CHARMM Drude simulations exhibits satisfactory convergence properties, as both error metrics fall below the $1\,k_BT$ threshold within the 5 ns equilibration period (Figure 5.13B–C). As in the case of the AMOEBA force field, the equilibration occurs more slowly than with the additive CHARMM36m force field, supporting the idea that polarisable ions have a stronger tendency to associate with the pore wall. Unfortunately, the umbrella histogram (Figure 5.13A) exhibits a high degree of variability, so that despite the good convergence properties it is not certain that the corresponding PMF is fully reliable.

| Property | Units | AMOEBA14 | SWM4-NDP | mTIP3P | TIP4P/2005 |
|---|---|---|---|---|---|
| $r_1$ | nm | 0.255 | 0.237 | 0.235 | 0.239 |
| $R_1$ | nm | 0.331 | 0.306 | 0.316 | 0.319 |
| $N_1$ | — | 6.0 | 5.4 | 5.8 | 5.9 |
| $r_2$ | nm | 0.450 | 0.454 | 0.468 | 0.449 |
| $R_2$ | nm | 0.544 | 0.561 | 0.577 | 0.547 |
| $N_2$ | — | 16.7 | 18.7 | 20.7 | 17.8 |

**Table 5.1:** Bulk hydration properties of $Na^+$ ions. The location of the $i$-th peak, $r_i$, and $i$-th minimum, $R_i$, of the sodium-oxygen RDF are given. The table also lists the corresponding coordination number for the first two hydration shells, $N_i$, i.e. the average number of water molecules contained within the first two peaks of the RDF.

### Ion Hydration

Having discussed the potential of mean force profile of ions permeating the 5-$HT_3$ receptor, it is worthwhile to consider whether the ions remain hydrated as they pass through the channel. This section will first address the structure of the hydration shell of ions solvated in bulk water before proceeding to investigate how the hydration number of $Na^+$ and $Cl^-$ ions varies as they are restrained at subsequent positions along the channel pore.

As can be seen from Figure 5.14A, the RDF for water around a $Na^+$ ion exhibits a sharp first peak at a distance of ∼0.25 nm and a somewhat broader second peak at a distance of ∼0.45 nm. These two peaks are clearly separated and the RDF approaches zero in its first minimum at a distance of ∼0.3 nm, thus clearly defining the first hydration shell. The next minimum, delineating the extent of the second hydration shell, is located at ∼0.55 nm. Beyond the second peak, the RDF is close to unity and the exact location of further minima and maxima cannot be quantified accurately.

The location of the first two minima and maxima of the RDF for each water model are listed in Table 5.1 alongside the number of water molecules in the first and second hydration shells. The AMOEBA14 model predicts the largest first shell radius and the

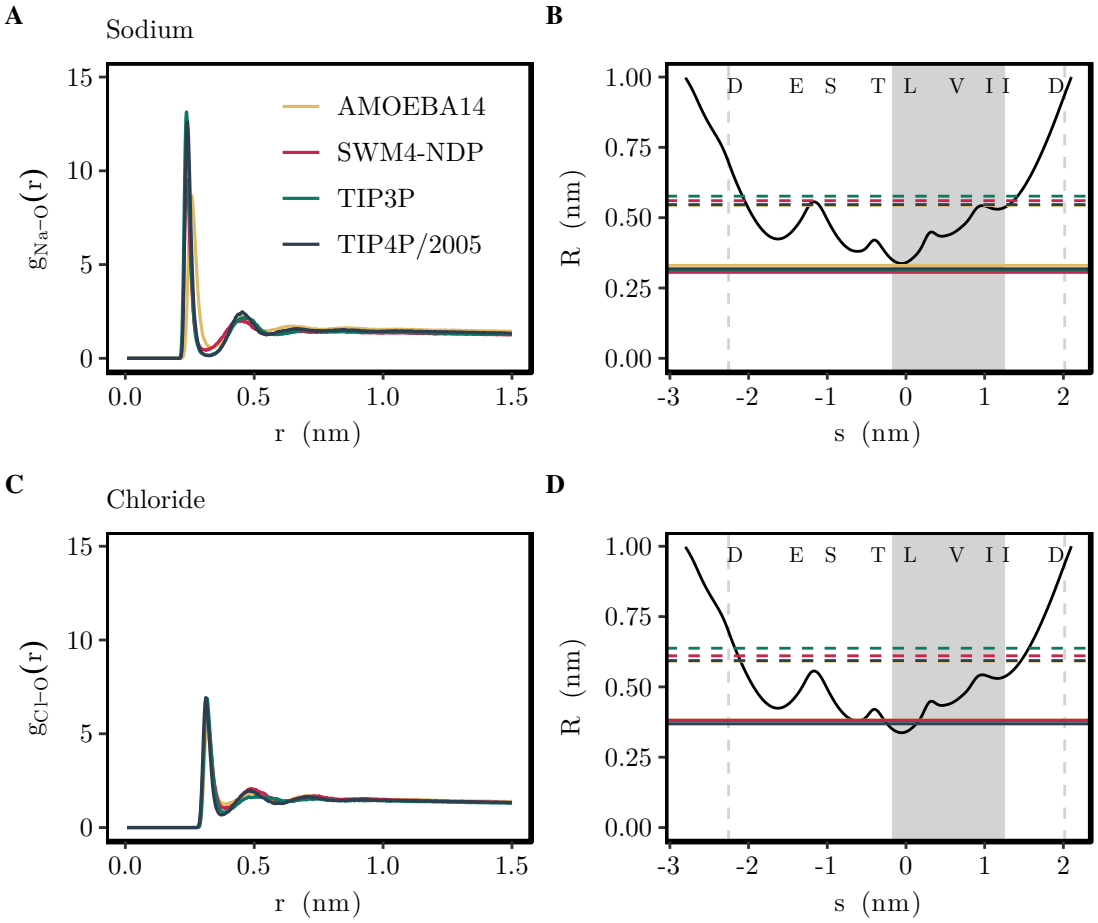



**Figure 5.14:** Ionic hydration shells for different water models. (A,C) Radial distribution function of water oxygens around (A) $Na^+$ ions and (C) $Cl^-$ ions in bulk water. The radial distribution function was calculated for an ion located in the bulk water regime outside the channel pore and is normalised to take a value of 1 at large separations. (B,D) Radius of the first (solid lines) and second (dashed lines) hydration shell of (B) $Na^+$ ions and (D) $Cl^-$ ions compared to the radius profile of the open state (PDB ID: 6DG8) structure of the 5-$HT_3$ receptor. Hydration shell radii were calculated from the minima of the radial distribution function. The shaded area and vertical dashed lines represent the hydrophobic gate region and the extent of the protein respectively.

SWM4-NDP model predicts the smallest, while the additive mTIP3P and TIP4P/2005 models yield intermediate values. Reassuringly, the location of the first peak agrees reasonably well with the mean sodium-oxygen distance of $(0.243 \pm 0.002)$ nm obtained from large-angle X-ray scattering (LAXS) experiments[272] as well as with ab initio MD calculations,[308] which yield a first peak distance in the range from 0.241 nm to 0.255 nm

depending on the choice of basis set. Corresponding to the size of the first hydration shell, the AMOEBA14 model predicts a coordination number of 6.0, whereas the SWM4-NDP model yields a value of only 5.4. The mTIP3P and TIP4P/2005 models predict a hydration number of 5.8 and 5.9 respectively.

A hydration number of 6.0 has previously been reported for the AMOEBA03 model,[187] while an earlier study employing the SWM4-NDP model[300] obtained a slightly higher value of 5.6. By comparison, predictions from recent ab initio MD simulations[308,309] fall in the range from 5.1 to 6.1. Historic experiments[310] place the hydration number of $Na^+$ ions in the range from 4.0 to 8.0, while recent LAXS experiments[272] favour a value of 6.0. To the best of the author's knowledge, no experimental reference data characterising the size and occupancy of the second hydration shell is available.

As shown in Figure 5.14B, the first hydration shell of a $Na^+$ ion is smaller than the narrowest constriction of the 5-$HT_3$ receptor pore in the open state (PDB ID: 6DG8). By contrast, the radius of the second hydration shell exceeds the pore radius nearly everywhere. Consequently, $Na^+$ ions can remain partially hydrated as they pass through the channel, but are expected to be stripped of part or all of the water molecules in the second hydration shell due to steric interactions with the pore wall.

This expectation is largely confirmed by the hydration number profiles shown in Figure 5.15. The CHARMM Drude results are not included in this figure, as the corresponding PMF profile is not converged (Figure 5.13). For the mTIP3P, TIP4P/2005, and AMOEBA14 water models, the first shell hydration number decreases by no more than $\sim$0.5 compared to the respective bulk value when the ion is located in the hydrophobic gate region (Figure 5.15A). This small decrease can be explained by the fact that the ion is occasionally located near the pore wall and is consistent with an otherwise intact inner hydration shell. Contrastingly, the second shell hydration number profile exhibits a marked decrease as the distance to the bulk water regime increases (Figure 5.15B). However, near the $-1'$ glutamic and $20'$ aspartic acid residues, both the first and

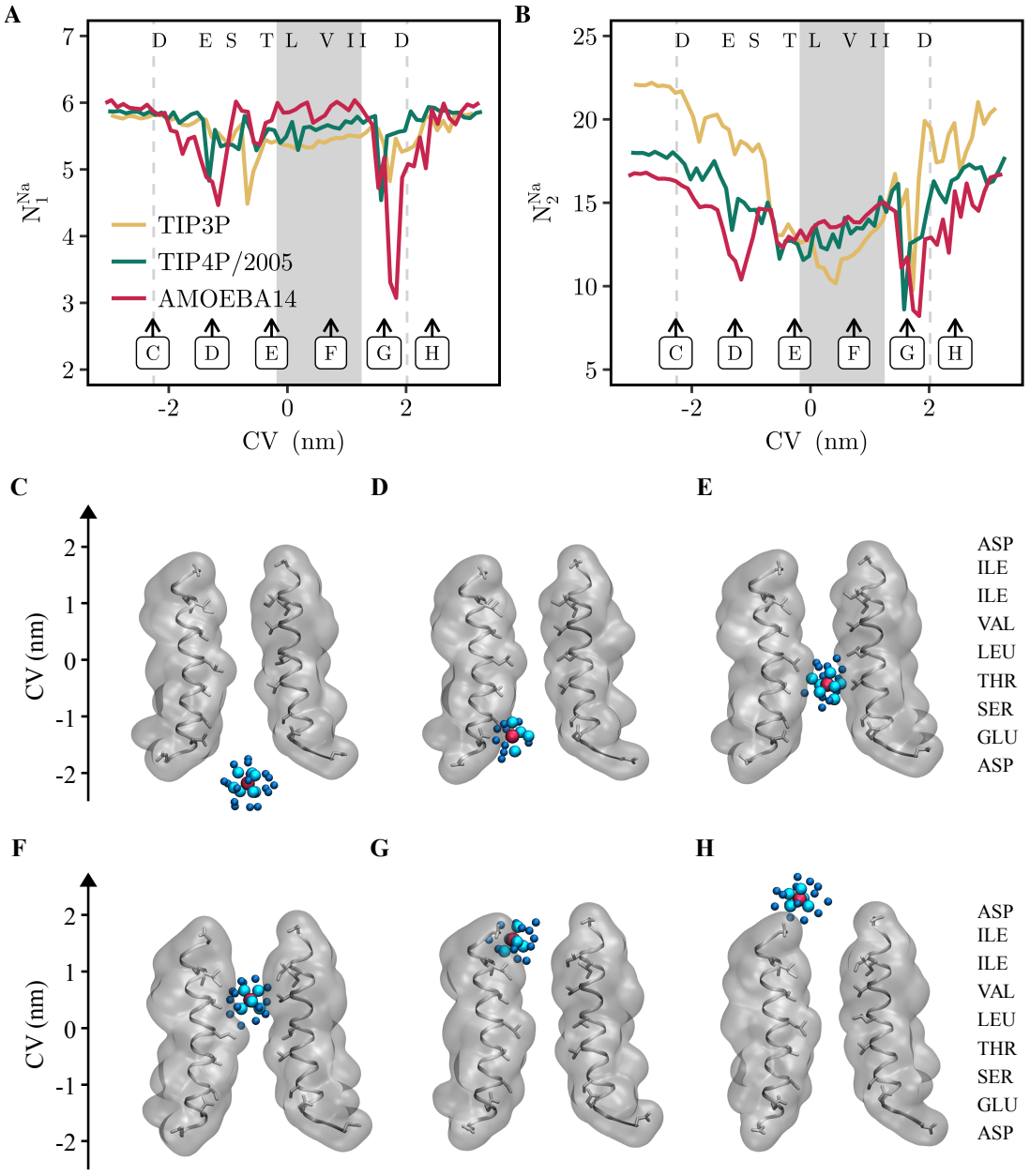



**Figure 5.15:** Hydration number profile for $Na^+$ ions. (A – B) Variation of the first shell (A) and second shell (B) hydration number of $Na^+$ ions along the pore of the open conformation 5-$HT_3$ receptor (PDB ID: 6DG8). The shaded area and vertical dashed lines represent the hydrophobic gate region and the extent of the protein respectively. The position of the snapshots shown in panels C through H is indicated by labelled arrows. (C – H) Representative snapshots from six umbrella sampling windows employing the AMOEBA force field. The restrained ion is shown as red van der Waals sphere, while oxygens of water molecules in the first and second hydration shell are shown in cyan and blue respectively. Two of the five pore-lining protein helices are shown in both cartoon and surface representation.

| Property | Units | AMOEBA14 | SWM4-NDP | mTIP3P | TIP4P/2005 |
|---|---|---|---|---|---|
| $r_1$ | nm | 0.317 | 0.315 | 0.314 | 0.309 |
| $R_1$ | nm | 0.383 | 0.382 | 0.379 | 0.369 |
| $N_1$ | — | 6.6 | 6.3 | 7.4 | 6.6 |
| $r_2$ | nm | 0.514 | 0.482 | 0.502 | 0.484 |
| $R_2$ | nm | 0.592 | 0.611 | 0.638 | 0.594 |
| $N_2$ | — | 20.2 | 20.6 | 26.1 | 20.5 |

**Table 5.2:** Bulk hydration properties of $Cl^-$ ions. The location of the $i$-th peak, $r_i$, and $i$-th minimum, $R_i$, of the chloride-oxygen RDF are given. The table also lists the corresponding coordination number for the first two hydration shells, $N_i$, i.e. the average number of water molecules contained within the first two peaks of the RDF.

second shell hydration number for the AMOEBA model is reduced by ∼1.5 and ∼2.5 respectively (Figure 5.15A–B). Similar but smaller decreases can also be observed for the mTIP3P and TIP4P/2005 models.

This can be explained through the interactions of $Na^+$ with one of the negatively charged side chains which attracts the ion into a position close to the pore surface. This in turn displaces a significant fraction of the first shell water molecules, as is illustrated in the simulation snapshots shown in Figure 5.15C–H. In the vicinity of uncharged sidechains (Figures 5.15E and 5.15F), $Na^+$ is mostly located in the centre of the pore and maintains a hydration shell similar to an ion placed in bulk water (Figures 5.15C and 5.15H). However, near charged side chains the ion is typically located much closer to the pore wall (Figures 5.15D and 5.15G). While this effect appears to exist for the additive mTIP3P and TIP4P/2005 water molecules, it is significantly enhanced for the polarisable AMOEBA force field.

The distribution of water around $Cl^-$ ions is shown in Figure 5.14C and the corresponding shell radii and hydration numbers for each water model are summarised in Table 5.2. Compared to $Na^+$, $Cl^-$ ions maintain a larger distance from the water oxygens, which manifests itself in an chloride-oxygen RDF that is shifted towards greater

separations (Figure 5.14C). This is caused in part by the larger van der Waals radius of $Cl^-$, but is also due to the fact that, on average, the water molecules orient themselves in such a way that the electronegative oxygen atom points away from the $Cl^-$ ion. All four water models yield similar values for the location of the first RDF peak and the radius of the first hydration shell. The predicted chloride-oxygen distances are also in reasonable agreement with the results of ab initio calculations[308] ranging from 0.317 nm to 0.319 nm, as well as with the experimental value of $(0.31 \pm 0.01)$ nm obtained from neutron diffraction data.[311]

However, the predicted first shell hydration number varies from 6.3 for the SWM4-NDP model to as much as 7.4 for the mTIP3P model, while both the AMOEBA14 and TIP4P/2005 models yield a value of 6.6. This exceeds the previously reported value of 6.0 for the AMOEBA03 water model,[187] but is reasonably close to the value of 6.5 reported for the SWM4-NDP model.[300] All models except mTIP3P also fall within the range of 6.3 to 6.6 predicted by ab initio calculations.[308] However, neutron scattering data[311] suggests a hydration number of $7.0 \pm 0.4$, although earlier experiments[312] favour a value between 5.3 and 6.2. The second shell hydration number is predicted to be ∼20.5 by all water models except mTIP3P, which yields a significantly higher value of 26.1. Unfortunately, as in the case of $Na^+$, no experimental reference data for the structure of the second hydration shell is available.

In Figure 5.14D, the radius of a hydrated $Cl^-$ ion is compared to the size of the pore in the open state structure of the 5-$HT_3$ receptor (PDB ID: 6DG8). In contrast to $Na^+$ and irrespective of the choice of water model, the first hydration shell of $Cl^-$ cannot be accommodated near the narrowest constriction of the pore caused by the 9′ leucine residues. However, this steric effect is likely only small and throughout most of the pore the first hydration shell can remain intact. The second hydration shell on the other hand cannot be accommodated in the pore, regardless of which water model is used.

As can be seen in Figure 5.16A, there is a clear difference in how the first shell hy-

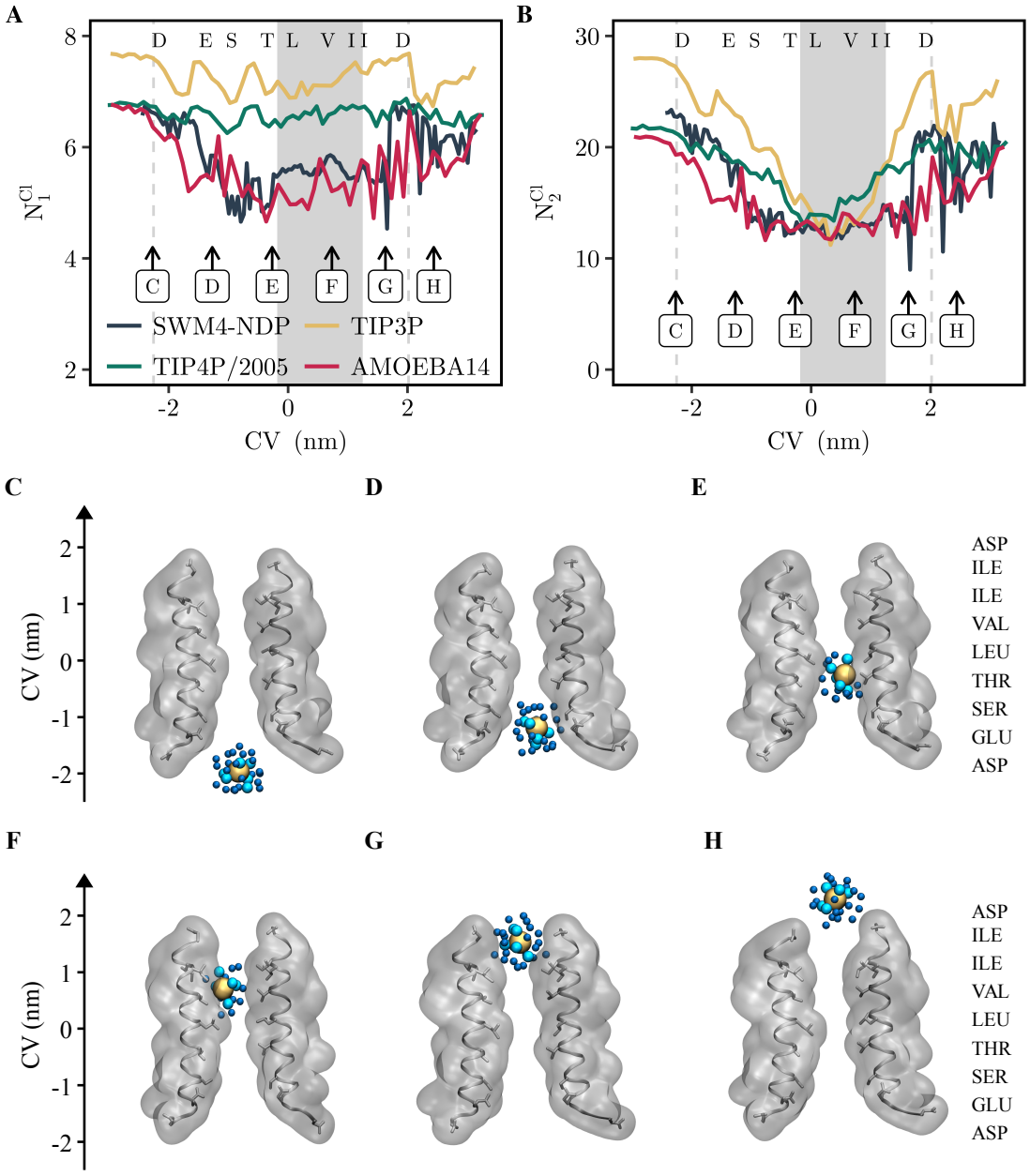



**Figure 5.16:** Hydration number profile for $Cl^-$ ions. (A – B) Variation of the first shell (A) and second shell (B) hydration number of $Cl^-$ ions along the pore of the open conformation 5-$HT_3$-receptor structure (PDB ID: 6DG8). The shaded area and vertical dashed lines represent the hydrophobic gate region and the extent of the protein respectively. The position of the snapshots shown in panels C through H is indicated by labelled arrows. (C – H) Representative snapshots from six umbrella sampling windows employing the AMOEBA force field. The restrained ion is shown as yellow van der Waals sphere, while oxygens of water molecules in the first and second hydration shell are shown in cyan and blue respectively. Two of the five pore-lining protein helices are shown in both cartoon and surface representation.

dration number varies along the pore between the additive mTIP3P and TIP4P/2005 water models on the one hand and the polarisable AMOEBA14 and SWM4-NDP models on the other hand. With both additive models, the hydration number remains mostly constant along the entire length of the pore and only exhibits local fluctuations of a magnitude below ∼0.5. By contrast, the two polarisable models predict a marked decrease of the hydration number throughout the pore, where it falls to ∼5.0 from a bulk value of ∼6.5. This decrease is moderate near the charged residue at the pore openings and is most pronounced between the 2′ serine and 17′ isoleucine residues.

This behaviour suggests that in the polarisable cases the ion spends a larger amount of time close to the pore surface. When the ion is placed near uncharged side chains (Figure 5.16E–F), it induces a dipole moment on the protein atoms that in turn enables an attractive interaction between the ion and the pore wall. In the vicinity of the negatively charged aspartic and glutamic acid residues (Figures 5.16D and 5.16G), this effect appears to still be present, but is partly counteracted by charge-charge repulsion, so that the ion is located closer to the centre of the pore and remains more fully hydrated.

Polarisable $Cl^-$ and $Na^+$ ions thus exhibit a very different behaviour in the vicinity of uncharged side chains. While $Cl^-$ ions substitute part of their hydration shell for a tight association with the amino acids, $Na^+$ ions retain most of their hydration shell by maintaining a larger distance from the pore surface. However, $Na^+$ ions are likely able to generate induced dipoles of a magnitude comparable to those caused by $Cl^-$. The observed difference is therefore probably due to the tighter binding of hydration shell waters to $Na^+$, whose first hydration shell is smaller by ∼0.05 nm than that of $Cl^-$. Note that this effect cannot be reproduced by an additive force field.

Finally, Figure 5.16B shows that the second shell hydration number decreases markedly for all water models, consistent with the expectation from Figure 5.14D. In all cases, the minimum of ∼12.5 is reached in the middle of the pore, close to the 9′ leucine residue. However, for the polarisable models, the second hydration shell is depleted even fur-

ther out, in agreement with the hypothesis that polarisable $Cl^-$ ions have a tendency to associate with the pore wall. It is interesting to note that in a simulation study of polarisable ions near the air-water interface, $Cl^-$ exhibited a preference for positions in the interfacial region, an effect that could not be observed with a non-polarisable force field.[313]

## 5.3 Discussion

The present chapter set out to address the question whether polarisation effects are relevant in hydrophobic gating and this question can now be answered in the affirmative. As has been shown, there is a significant variation of the dipole moment of water as it passes through the de-wetted section of the 5-$HT_3$ receptor pore (Figure 5.11). This is observed with both the AMOEBA and CHARMM Drude force fields and is largely independent of the protein model and its lipid bilayer environment (Figure 5.10).

Previous studies have investigated the behaviour of polarisable water near protein surfaces and it was found that the dipole moment of SWM4-NDP water decreases by no more than 0.3 D in a 0.5 nm thick surface layer around solvated model peptides.[285] Furthermore, the average dipole moment of SWM4-NDP water in the vicinity of (charged) glutamate residues has been reported to increase by ∼0.1 D compared to bulk water.[160] In comparison to this, the SWM4-NDP model as well as both AMOEBA models experience a remarkable reduction of ∼0.5 D in the hydrophobic gate of the 5-$HT_3$ receptor.

Two independent models for induced polarisation thus consistently predict that inside an ion channel pore the molecular dipole moment of water undergoes a significant change and may approach a vapour-like value in a hydrophobic gate. Considering the success of these models in describing the interface properties of bulk water (Figure 5.1A) and the importance of interfacial tensions in the theory of hydrophobic gating,[44] it seems likely that reliable quantitative estimates for the free energy of hydrating an ion channel pore can only be obtained when polarisation is modelled explicitly.

In the closed and intermediate conformations of the 5-$HT_3$ receptor, simulations employing polarisable force fields result in an almost fully de-wetted pore (Figure 5.5), consistent with the predictions made by a variety of additive water models (Figure 5.6). However, the fact that in some of these cases no transitions between the liquid and vapour states occur during the entire simulation implies that for these structures a quantitative comparison is not possible based on simple equilibrium simulations. Instead, a grand canonical biasing method would be required.

By contrast, the pore radius and hydrophobicity of the open conformation 5-$HT_3$ receptor create an environment in which neither the liquid nor the vapour state are strongly favoured. In this case, predictions made by different water models disagree significantly (Figure 5.7A). Moreover, there is also no agreement between the polarisable CHARMM Drude and AMOEBA force fields, although both AMOEBA water models yield similar results.

Since there is no experimental data on the hydration equilibrium of ion channel pores, it cannot directly be assessed which water model makes the most accurate prediction. An attempt was therefore made to relate the free energy of pore hydration to the liquid-vapour surface tension of bulk water, for which experimental reference values are available (Figure 5.7A). Unfortunately, no clear trend emerged, which suggests that in addition to the properties of the water model itself, its interaction with the channel protein are crucial.

Consequently, the most accurate results can be expected to come from simulations employing a water model that accurately reproduces the properties of bulk water in conjunction with a consistently parameterised protein force field. Future work should therefore evaluate novel additive force fields like AMBER-FB15[314] or AMBER ff15iqp,[315] which respectively use TIP3P-FB[201] and SPC/$E_b$[316] as their host water models. While these models do not achieve the accuracy of TIP4P/2005 and OPC, they outperform traditional water models like TIP3P and offer superior performance compared to polarisable

force fields.

It is also worth noting that although polarisable models in principle offer the advantage of transferability, there is no guarantee that their current parameterisations provide a more accurate description of physical reality than the most recent additive force fields. For example, in the SAMPL4 hydration free energy prediction challenge,[317] the AMOEBA force field performed significantly worse than the additive generalised AMBER force field (GAFF).[318] Polarisable force fields may thus need further refinements before surpassing their additive counterparts, which have already undergone a long history of improvements.

What do these results imply for the functional annotation of ion channels based on equilibrium MD simulations? Firstly, it is reassuring that pore de-wetting can still be observed with the AMOEBA and CHARMM Drude force fields, as this rules out that induced polarisation qualitatively changes the collective behaviour of nanoconfined water. Secondly, the results suggest that the choice of water model can in some cases influence the predicted hydration state of the channel pore. Fortunately, it should be possible to identify these cases from simulations with a single water model, as they will exhibit a significant number of transitions between the liquid and vapour states. It is then possible to subject these channel structures to more detailed computational analyses.

Umbrella sampling simulations employing a polarisable force field have thus far only been reported for one ion channel system. Free energy profiles for $Na^+$ and $K^+$ ions in the gramicidin A channel were calculated with the AMOEBA force field and it was concluded that the explicit inclusion of polarisation effects significantly reduces the energetic barrier experienced by both ion species compared to simulations with the CHARMM27 force field.[101] Based on a generalised Langevin framework, Peng et al. [101] also derived an estimate for the conductance of gramicidin A from the ionic velocity autocorrelation function, which is in better agreement with experiment than any previous uncorrected

simulation result employing an additive force field.

By contrast, in this work the AMOEBA force field predicts a higher free energy barrier than the CHARMM36m force field for both $Na^+$ and $Cl^-$. Moreover, AMOEBA simulations result in a more rugged free energy landscape than additive force fields, indicating that induced polarisation permits more specific interactions between the protein and the ions and gives rise to favourable binding poses that do not exist in an additive force field. Umbrella sampling simulations with the CHARMM Drude force field exhibit poor convergence properties, but at least for $Cl^-$ ions the resulting free energy profile is also characterised by a high degree of ruggedness.

Unfortunately, the generalised Langevin equation cannot be used to estimate channel conductance from the umbrella sampling simulations presented here. This is because the Andersen and dual-Langevin thermostats employed in the AMOEBA and CHARMM Drude simulations respectively regulate temperature by randomising and thus decorrelating velocities. Nonetheless, the question of how polarisation influences the dynamics of ion conduction is important. The investigation of ion hydration conducted here suggests that with a polarisable force field, ions interact with the pore wall more favourably, which would imply a higher degree of friction and thus a lower channel conductivity. Future studies could address this e.g. by directly monitoring the flow of ions through the channel by means of computational electrophysiology simulations.[93]

## 5.4 Summary and Conclusions

It is the highly polar nature of the water molecule that gives rise to the expulsion of water from nanopores with hydrophobic surfaces and thus to the hydrophobic gating mechanism in biological ion channels. This chapter has explored the importance of induced polarisation for this mechanism, the current understanding of which has so far been based primarily on MD simulations using pairwise additive force fields. To this end, the next generation AMOEBA and CHARMM Drude force fields were employed,

which represent two distinct approaches of incorporating atomic polarisation into the MD framework. Protocols for polarisable MD simulations of membrane proteins were implemented in OpenMM and validated through an analysis of the equilibration of their thermodynamic state variables.

Induced polarisation was indeed found to play an important role in the modelling of nanoconfined water. This is exemplified by the variation of its molecular dipole moment in the pore of the 5-$HT_3$ receptor. While largely bulk-like throughout most of the channel, the dipole moment decreases appreciably in the hydrophobic gate, where it assumes a value close to that of gas phase water. Moreover, a comparison of several additive and polarisable water models found them to consistently predict de-wetting in the closed and intermediate state structures of the 5-$HT_3$ receptor. However, in the open conformation of the channel, where the free energy difference between the liquid and vapour states is close to zero, predictions can vary considerably. In this case, an accurate prediction of the hydration state of the pore necessitates not only a water model whose bulk properties are in quantitative agreement with experiment, but also a correct description of surface tension at the water-protein interface.

Beyond water, the influence of induced polarisation on ion conduction was studied by means of umbrella sampling. It was found that polarisable force fields yield richer free energy landscapes that are less smooth than those obtained with additive force fields. At the same time, explicitly modelling polarisation allows ions to interact more intensively with the pore wall. Polarisable $Na^+$ ions bind more tightly to charged amino acids than they would with additive force fields, while polarisable $Cl^-$ ions exhibit a high affinity to uncharged residues. It is hypothesised that this increases the friction between the ions and the channel protein and could slow down the motion of ions through the channel pore.

Clearly, extensions of classical MD capable of describing the effects of electronic polarisation have an important role to play in future studies of biomolecular transport

processes. However, compared to the traditional all-atom approach, polarisable force fields incur a significant additional computational cost and have not yet undergone a significant amount of testing on complex systems and over long time scales. For the foreseeable future, additive force fields will therefore remain a cornerstone of the high-throughput functional annotation procedure pursued in this thesis.

This is what the world is for —
making electricity.

*(MGMT)*

# 6

# Electrowetting of Hydrophobic Gates

In Chapters 3 and 4, equilibrium MD simulations were used to investigate the behaviour of water confined in ion channel pores. It was shown that hydrophobic amino acid side chains may induce liquid-vapour phase transitions and that particularly hydrophobic pores may permanently enter a vapour state devoid of water. This pore de-wetting consequently creates a free energy barrier for ion permeation, as was shown in Chapter 5 through umbrella sampling. However, biological ion channels frequently experience a transmembrane voltage, which has a magnitude of $\sim$100 mV under typical physiological conditions in mammalian cells and may reach up to $\sim$200 mV in bacteria. The effects of such a transmembrane potential were absent from the simulations presented thus far and their significance for the functional annotation of ion channel structures through equilibrium MD is the topic of this chapter.

It was originally demonstrated in simulations of cylindrical model nanopores that the application of an electric field can hydrate an otherwise de-wetted hydrophobic constriction, which is thus rendered ion permeable.[47] This is illustrated in Figure 6.1A. Subsequently, Dzubiella and Hansen [48] showed that this electrowetting effect can be modelled mathematically by incorporating the electrical energy contained in a volume element permeated by an electric field into the thermodynamic model of Beckstein and

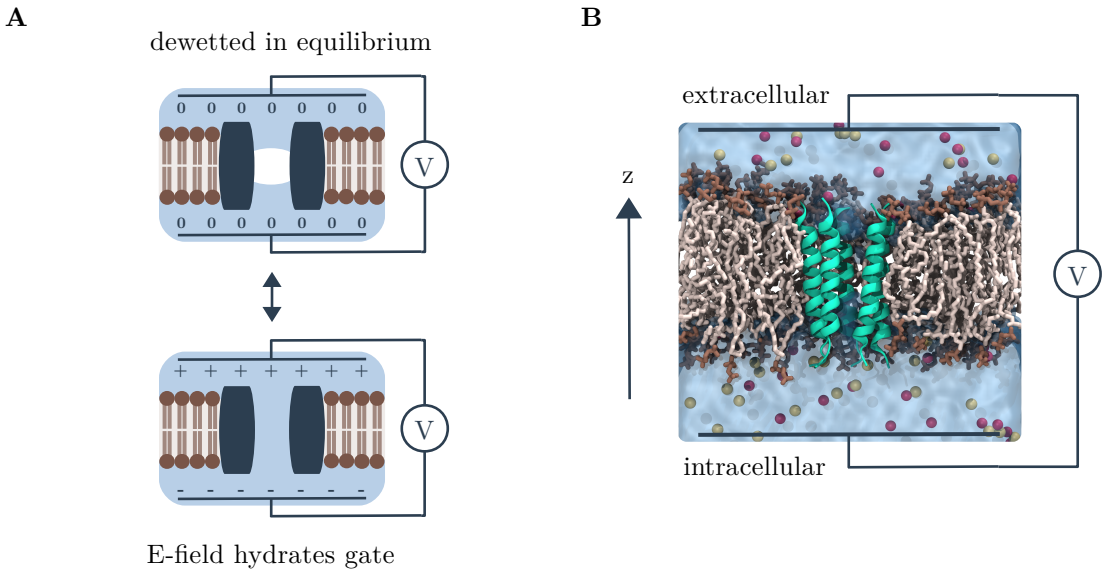



**Figure 6.1:** Electrowetting of hydrophobic gates. (A) Schematic representation of electrowetting. In the absence of an electric field, a hydrophobic nanopore exists in the vapour state. When a transmembrane voltage is applied, the energy contained in the electric field causes hydration of the nanopore. (B) MD simulation setup. The M2 helix bundle of the 5-$HT_3$ receptor (green cartoon) is embedded in a lipid bilayer (brown licorice). Water is represented as a transparent surface and $Na^+$ and $Cl^-$ ions are shown as red and yellow van der Waals spheres respectively. The effect of a transmembrane potential is modelled by applying a constant electric field to all atomic charges in the system.

Sansom [44] (see Section 1.2.4 for details). However, the authors also pointed out that an electric field causes a significant distortion of the hydrogen bonding network of water and noted that this effect would also need to be taken into account to accurately describe electrowetting.[48]

Indeed, Daub et al. [319] conducted simulations of water nanodroplets on apolar surfaces and demonstrated that an electric field modifies the interfacial tensions (and hence the contact angle) by influencing both the orientation and hydrogen bonding structure of water molecules located on the droplet surface. Similarly, in MD simulations of water in planar confinement, Bratko et al. [320] observed decreasing interfacial tensions under the influence of an electric field as well as a field-induced change in the average number of hydrogen bonds of interfacial water molecules. Consequently, both studies[319,320] also find an influence of the direction of the electric field that cannot be accounted for by the

thermodynamic model of Dzubiella and Hansen [48].

Relatively little work has thus far been conducted to quantify the effect of electrowetting in biological ion channels. Spronk et al. [61] performed MD simulations of the mechanosensitive channel of small conductance (MscS) and found that its hydrophobic gate could be partially hydrated when a transmembrane voltage of 0.22 V was applied, while a voltage of 1.1 V was sufficient to fully hydrate the channel. These results were corroborated by Sotomayor et al. [321], who also observed hydration of the hydrophobic gate in MscS with a transmembrane potential of 1.2 V. More recently, using computational electrophysiology,[93] Trick et al. [79] demonstrated that a hydrophobic gate engineered into biomimetic nanopores can be hydrated by voltages ranging from 0.8 V to 1.2 V. In addition, electrowetting has also been reported for carbon nanotubes.[86]

Experimentally, electrowetting has been observed in several investigations of artificial nanopores.[322] Smirnov et al. [323] showed that hydrophobic pores in silicon nitride membranes exhibit voltage-gating behaviour, when an external electric field is used to bias the water confined within these pores towards the liquid state. Similarly, it was demonstrated by Powell et al. [83] that nanopores track-etched into polyethylene terephthalate (PET) membranes can be rendered functional through the application of an electric field. Xiao et al. [324] confirmed this result and demonstrated that the threshold voltage required for electrowetting of PET nanopores can be controlled through its pH-dependence.

It is thus clear that electrowetting can in principle open a hydrophobic gate, which raises the question whether the predictions made previously are robust with respect to the influence of a physiological electric field. While existing studies of the MscS channel[61,321] suggest that electrowetting might require supraphysiological transmembrane voltages, the underlying simulations were relatively short ($\sim$10 ns). Moreover, the exact relationship between pore hydration probability and the magnitude of an applied electric field has never been determined quantitatively for a biological ion channel.

These issues will be addressed in this chapter. Using the 5-$HT_3$ receptor as a test system, the field-induced hydration of its pore will be analysed through MD simulations and the critical voltage required for opening its hydrophobic gate will be determined in the context of thermodynamic theory. It will be shown that pore de-wetting effectively prevents ion conduction even at large electric fields.

Subsequently, the interplay between a transmembrane voltage and the electric field generated by the charged residues of the ion channel will be investigated through continuum electrostatics calculations. Moreover, electric-field-induced pore wetting will be used for a detailed study of the kinetics of liquid-vapour transitions. From this analysis, a minimum simulation time required for functionally annotating ion channel structures through equilibrium MD simulations will be derived.

Before turning to these results in Section 6.2, details of the underlying simulation protocols and data analysis will be discussed in Section 6.1. A discussion of the results can be found in Section 6.3 and the main findings of this chapter are summarised in Section 6.4.

## 6.1 Methods

The 5-$HT_3$ receptor introduced in Chapter 4 and further discussed in Chapter 5 is used here as a test system to probe the influence of an external electric field on the hydration of a hydrophobic gate. This is primarily assessed through MD simulations as detailed in Section 6.1.1, as well as through a set of continuum electrostatics calculations described in Section 6.1.2. Details of data analysis and statistical modelling are described in Section 6.1.3.

### 6.1.1 Molecular Dynamics Simulation

Atomistic MD simulations employing a constant electric field were carried out for the closed (PDB ID: 4PIR) and open (PDB ID: 6DG8) state structures of the 5-$HT_3$ receptor.

All simulations were performed in GROMACS 2018[98] using the protocol described below.

### Structural Model and System Preparation

Structures of the 5-$HT_3$ receptor were obtained from the PDB and missing atoms were added using the WHAT IF tool.[242] Since it was shown in Section 4.2.1 that the hydration of a hydrophobic gate is not strongly influenced by distal protein domains, only the M2 helix bundle (residues 247 – 271) was considered here. A 100 ns long MD simulation of this M2 helix bundle together with DOPC lipids and water was then carried out using the MARTINI force field.[302] During this coarse-grained simulation lipid molecules self-assemble into a bilayer structure around the protein.

Following the self-assembly simulation, the protein-bilayer system was converted back to an atomistic representation and was solvated in a 150 mmol $L^{-1}$ NaCl solution. The system was then equilibrated through a 10 ns long MD simulation employing the mTIP3P water model[120] and the CHARMM36m force field.[109,303] During this equilibration period the simulation parameters were the same as those used for the productive simulations described below. However, no electric field was applied during the equilibration stage.

### Constant Electric Field Simulations

Productive simulations were carried out using the CHARMM36m protein force field[109] and associated lipid parameters[303] together with the mTIP3P,[120] SPC/E,[122] TIP4P,[118] and TIP4P/2005[123] water models. During these simulations, the effect of an external electric field, $\mathbf{E}$, was modelled through an additional force, $q_i\mathbf{E}$, acting on each atomic partial charge, $q_i$. The electric field was applied in the $z$-direction perpendicular to the plane of the membrane so that the field can be written as $\mathbf{E} = (0, 0, E)^\intercal$. Figure 6.1B illustrates this simulation setup.

It has been shown[92,325] that under periodic boundary conditions the application of an external field as described above is equivalent to the application of a transmembrane

voltage as illustrated in Figure 6.1B. It is important to note that the voltage across the membrane is given by $V = EL_z$, where $L_z$ is the length of the unit cell in the $z$-direction rather than the thickness of the membrane. For the system considered here $L_z \approx 8.5\,\mathrm{nm}$, although the exact size of the simulation cell varies due to pressure coupling.

The magnitude of the electric field, $E$, was varied in steps of $25\,\mathrm{mV\,nm^{-1}}$ between $-250\,\mathrm{mV\,nm^{-1}}$ and $250\,\mathrm{mV\,nm^{-1}}$ for the closed state structure (PDB ID: 4PIR) and between $-150\,\mathrm{mV\,nm^{-1}}$ and $150\,\mathrm{mV\,nm^{-1}}$ for the open state structure (PDB ID: 6DG8). In terms of transmembrane potential, this corresponds to a step size of 0.2125 V as well as maximum transmembrane voltages of $\pm 2.125\,\mathrm{V}$ and $\pm 1.275\,\mathrm{V}$ respectively. In each case, the range was chosen to ensure that a complete transition from a (partially) de-wetted to a fully hydrated pore could be observed.

Preliminary simulations were carried out to assess the stability of the simulation system under electric fields of this magnitude. It was found that electroporation,[326–328] i.e. the rupture of the lipid bilayer and the formation of a continuous water channel directly through the membrane, occurs for $|E| > 150\,\mathrm{mV\,nm^{-1}}$. In order to assess pore hydration properties under these high-field conditions, a harmonic restraining force of $1000\,\mathrm{kJ\,mol^{-1}\,nm^{-2}}$ was placed on the heavy atoms of all lipid molecules (after the initial 10 ns long equilibration period), which prevented electroporation.

### Simulation Details

Three independent simulations with a duration of 150 ns for the closed conformation (PDB ID: 4PIR) and 50 ns for the open conformation (PDB ID: 6DG8) were carried out at each value of the electric field. Time integration was performed using a leapfrog integrator with a step size of 2 fs. Covalent bonds to hydrogen atoms were constrained using the LINCS algorithm[139] and water model geometry was enforced using the SETTLE method.[137] The conformation of the protein was maintained close to its experimentally determined structure through the application of a harmonic restraining potential

of 1000 kJ mol$^{-1}$ nm$^{-2}$ to all C$\alpha$ atoms.

Long-range electrostatic interactions were treated through the smooth PME method[152] employing a real space cutoff of 1.0 nm and a Fourier spacing of 0.12 nm, with charge interpolation onto the grid through fourth order B-splines. Lennard-Jones interactions were switched off between 1.0 nm and 1.2 nm and a long-range dispersion correction was applied to energy and pressure. All simulations were carried out in the isothermal-isobaric (NPT) ensemble. Temperature was maintained at 310 K through a velocity rescaling thermostat[142] with a coupling constant of 0.1 ps, while a pressure of 1.0 bar was enforced through a semi-isotropic Parrinello-Rahman barostat[143] with a coupling constant of 1.0 ps and a compressibility of $4.5 \times 10^{-5}$ bar$^{-1}$.

### 6.1.2 Poisson-Boltzman Electrostatics

In addition to the MD simulations described in the previous section, the electrostatic potential, $\phi(\mathbf{r})$, around the 5-HT$_3$ receptor was calculated by numerically solving the non-linear Poisson-Boltzmann equation[212,213] using the Adaptive Poisson-Boltzmann Solver (APBS).[219] All Poisson-Boltzmann calculations were carried out on the M2 helix bundle of the closed conformation of the channel (PDB ID: 4PIR), both in the presence and absence of a DOPC lipid bilayer.

#### Poisson-Boltzmann Parameters

A multigrid finite differencing method[329] was employed, in which the nonlinear Poisson-Boltzmann equation was first solved on a coarse grid, using Debye-Hückel boundary conditions, and subsequently on a finer grid focussed on the channel protein, with boundary conditions taken from the coarse calculation. The coarse grid was 2.5 times as large as the bounding box around the M2 helix bundle and had a grid spacing of 0.17 nm, while the fine grid was 1.5 times as large as the bounding box and had a grid spacing of 0.10 nm. It was validated through preliminary calculations that the resulting

electrostatic potential did not critically depend on the size and resolution of the grid.

For consistency with the MD simulations charges were assigned to the atomic sites according to the CHARMM36m force field and were discretised onto the grid through cubic B-spline interpolation. An electric permittivity of 2 for the protein and 78.5 for the solvent were used and calculations were carried out at a temperature of 310 K. A 150 mmol $L^{-1}$ NaCl electrolyte was used as solvent and ionic Born radii of 0.1680 nm for $Na^+$ and 0.1937 nm for $Cl^-$ were taken from Rashin and Honig [330]. Additional calculations were carried out to probe the influence of the electrostatic potential on these physical parameters.

#### Electrostatic Potential Profile

A one-dimensional electrostatic potential profile, $\phi(s)$, was calculated from the three-dimensional potential field, $\phi(\mathbf{r})$, resulting from the Poisson-Boltzman calculation as follows. CHAP version 0.9.1[26] was used to calculate a pathway through the ion channel pore, yielding a centre line spline curve, $\mathbf{s}(s)$, which represents the likely path along which an ion would permeate the channel (Section 3.2.1). The potential field was then evaluated along the spline curve according to $\phi(s) = \phi(\mathbf{s}(s))$ and trilinear interpolation was used where the spline curve did not pass through the APBS grid points.

### 6.1.3 Analysis

Unless otherwise specified, MD trajectories were post-processed using MDAnalysis version 0.19.0,[307] while further data analysis and visualisation were performed through custom scripts written in the R language.[331] Molecular visualisations were created in VMD.[248]

### Beckstein Openness and Kinetic Constants

CHAP version 0.9.1[26] was used to calculate a water density profile, $n(s,t)$, along the channel pore and the minimal instantaneous water density was used to determine a time series of Beckstein openness[44] according to

$$\omega(t) = \begin{cases} 1, & \text{if } \min_s n(s,t) > n_{\text{thres}} \\ 0, & \text{otherwise} \end{cases}, \tag{6.1}$$

where the threshold density $n_{\text{thres}} = 7.6\,\text{nm}^{-3}$ is chosen to classify the state of the channel as open ($\omega(t) = 1$) or closed ($\omega(t) = 0$). This value will be shown to lead to a minimal misclassification rate in Section 6.2.1.

The time-averaged Beckstein openness,

$$\langle\omega\rangle = \frac{1}{T_{\text{sim}}} \int_0^{T_{\text{sim}}} \omega(t)\ dt, \tag{6.2}$$

where $T_{\text{sim}}$ denotes the duration of the simulation, was subsequently calculated as a measure for the probability of finding the channel pore in a hydrated state. A thermodynamic model predicting the dependence of $\langle\omega\rangle$ on the external electric field, $E$, is described in the following section. From the Beckstein openness, the free energy difference between the liquid and vapour states was estimated as

$$\Delta\Omega = \Omega_{\text{v}} - \Omega_{\text{l}} = -k_{\text{B}}T \ln\left(\frac{1 - \langle\omega\rangle}{\langle\omega\rangle}\right), \tag{6.3}$$

where $k_{\text{B}}$ is the Boltzmann constant and $T$ is temperature. Note that for Equation 6.3 to be meaningful, at least one liquid-vapour transition needs to be observed during the course of a simulation, so that $\langle\omega\rangle \neq 0$.

In addition, the mean survival times of the liquid state, $\tau_{\text{l}}$, and vapour state, $\tau_{\text{v}}$ were determined. These two constants are inversely proportional to the respective rate

constants, $k_\mathrm{l}$ and $k_\mathrm{v}$, and are related to Beckstein openness through

$$\frac{1-\langle\omega\rangle}{\langle\omega\rangle} = \frac{k_\mathrm{l}}{k_\mathrm{v}} = \frac{\tau_\mathrm{v}}{\tau_\mathrm{l}}. \tag{6.4}$$

Survival times were estimated directly from the time series of $\omega(t)$ as the average duration of the liquid ($\omega(t) = 1$) and vapour states ($\omega(t) = 0$) respectively. The initial and final periods of the time series were excluded from this calculation, as it is unclear for how long beyond the end of a simulation a state would have persisted (or for how long before the beginning of a simulation it can be assumed to have persisted). Note that this implies that at least two liquid-vapour transitions need to be observed during a trajectory in order to estimate survival times. Furthermore, no time scales longer than the duration of a simulation (150 ns for simulations of the closed (PDB ID: 4PIR) and 50 ns for the open (PDB ID: 6DG8) conformation of the 5-$HT_3$ receptor) or shorter than the sampling time (0.1 ns for all simulations) could be estimated.

#### Bayesian Nonlinear Multilevel Model

As discussed in Section 1.2.4, the electrowetting of hydrophobic gates can be understood by considering the influence of an external electric field, $E$, on the free energy difference between the liquid and vapour states of the pore (Equation 1.5). Consequently, the expected Beckstein openness can be written as

$$\langle\omega\rangle = \left[1 + \exp\left(\frac{-\Delta\Omega_0 - m(E - E_\mathrm{int})^2}{k_\mathrm{B}T}\right)\right]^{-1}, \tag{6.5}$$

which constitutes a nonlinear model with three free parameters. Here $\Delta\Omega_0$ denotes the free energy difference in the absence of an external electric field, while $m$ represents how strongly the probability of hydration is coupled to the electric field. From Equation 1.5 it can be seen that this coupling strength is proportional to both the volume of the pore and the electric permittivity of the solvent. The third parameter, $E_\mathrm{int}$, is added

here to account for the possibility that an intrinsic electric field, created by an unequal distribution of charged amino acids along the pore, may add to or counteract the effect of the externally applied field.

A fit of Equation 6.5 to the observed Beckstein openness was obtained within the framework of Bayesian multilevel modelling.[332] Both $\Delta\Omega_0$ and $m$ were treated as dependent on both channel structure and water model, while $E_{\text{int}}$ was considered to only depend on the channel structure. The zero-field free energy difference, $\Delta\Omega_0$, may take both negative and positive values and a Gaussian prior with mean $\mu = 0$ and standard deviation $\sigma = 10\,k_{\text{B}}T$ was used. In contrast to this, the coupling strength, $m$, is positive definite by definition. A gamma prior with shape $\alpha = 2$ and rate $\beta = 0.01$ was chosen to enforce this constraint and sampling for this parameter was limited to positive values. While the intrinsic electric field, $E_{\text{int}}$, can in principle be both positive and negative, allowing this parameter to vary over the entire real range caused divergent transitions in the sampling process. Sampling for this parameter was therefore restricted to positive values (in accordance with the observed data) and a gamma prior with shape $\alpha = 2$ and rate $\beta = 100$ was chosen.

The model was implemented in the probabilistic programming language STAN[333] through the BRMS[334] library in R. Posterior samples were generated through a Hamiltonian Monte Carlo[335,336] procedure using the No-U-Turn sampler (NUTS).[337] For each of ten independent Markov Chains, 10 000 warm-up and 10 000 sampling iterations were carried out. The NUTS sampler used a target average acceptance ratio of 0.9 and a maximum tree depth of 15.

### Side Chain Orientation

To investigate the influence of the electric field on the amino acid side chains, their average orientation relative to the transmembrane axis was determined. This was quantified in terms of the polar angle, $\theta$, between the transmembrane axis and the vector pointing

from the C$\alpha$ atom of a residue to the centre of geometry of its side chain atoms. Thus, a residue with $\cos\theta = 1$ points up towards the extracellular domain, a residue with $\cos\theta = -1$ points down towards the intracellular domain, and a residue with $\cos\theta = 0$ lies in the plane defined by the lipid bilayer.

### Ion Transport and Electric Currents

Ion transport was analysed through custom scripts based on the MDAnalysis library.[307] For each value of the electric field, $E$, the net current, $I$, through the ion channel pore was calculated using the algorithm detailed below. The simulation cell was divided into three distinct domains representing the intra- and extracellular spaces as well as the ion channel pore. If an ion was located in a cylindrical region centred on the protein, it was considered to be inside the pore and was assigned the domain index $\gamma = 0$. An ion located outside this domain was assigned to the extracellular domain ($\gamma = +1$), if its $z$-coordinate placed it above the middle of the lipid bilayer, and to the intracellular domain ($\gamma = -1$) otherwise.

Over the course of a simulation, an ion may jump between these three domains and conduction events can be identified from the sequence of domain indices. In order to move directly between the extra- and intracellular domains ($\gamma = +1 \rightarrow \gamma = -1$ or $\gamma = -1 \rightarrow \gamma = +1$), an ion would have to either cross the lipid bilayer or transition through the periodic boundary. These events were disregarded and only true pore crossing events ($-1 \rightarrow 0 \rightarrow +1$ or $+1 \rightarrow 0 \rightarrow -1$) were counted towards the amount of charge conducted by the ion channel. Each such conduction event was assigned a crossing number, $\eta$, based on whether the particle moved from the intra- to the extracellular domain (i.e. in positive $z$-direction, $\eta = +1$) or vice versa ($\eta = -1$).

If an ion entered the pore ($\gamma = \pm 1 \rightarrow \gamma = 0$) at time $t^{\text{in}}$ and left it ($\gamma = 0 \rightarrow \gamma = \pm 1$) at time $t^{\text{out}}$, the time at which a crossing event occurred was taken to be $t^{\text{cross}} = (t^{\text{out}} + t^{\text{in}})/2$. Correspondingly, the ion dwell time inside the pore is given by

$T_{\text{pore}} = t^{\text{out}} - t^{\text{in}}$. Each ion may cross the pore multiple times and thus $t_{i,j}^{\text{cross}}$ denotes the time at which the $i$-th ion crossed the pore for the $j$-th time, while $M_i$ denotes the overall number of crossings of ion $i$. The cumulative amount of charge transported through the channel pore between the beginning of the simulation and time $t$ was then calculated according to

$$Q(t) = \sum_{i=1}^{N} \sum_{j=1}^{M_i} \int_0^t q_i \eta_{i,j} \delta(t' - t_{i,j}^{\text{cross}}) \, dt', \tag{6.6}$$

where $q_i$ is the charge of the $i$-th ion, $N$ is the total number of ions, and $\delta(t)$ is the Dirac delta function. Note that Equation 6.6 implies $Q(0) = 0$.

As will be shown, the rate of ion transport is approximately constant and thus $Q(t)$ grows linearly over time. A linear model of the form $Q(t) = It$ was therefore fit to the data using R,[338] where the proportionality constant $I$ represents the electric current through the channel pore. The individual contribution of $Na^+$ and $Cl^-$ ions to the net current was evaluated by carrying out the outer sum in Equation 6.6 over only a single ion species.

## 6.2 Results

### 6.2.1 Liquid-Vapour Equilibrium Under an Electric Field

Figure 6.2 exemplifies the electrowetting effect on the closed state structure of the 5-$HT_3$ receptor (PDB ID: 4PIR). Note that in this Figure, as throughout the entire results section, the electric field is consistently quantified in units of field strength (i.e. $\text{mV nm}^{-1}$) rather than in units of the corresponding transmembrane potential (i.e. V). Without an electric field, this channel remains de-wetted throughout most of a 150 ns long simulation and only enters the liquid state for a small number of very brief periods (Figure 6.2A). Correspondingly, the time-averaged water density depicted in Figure 6.2B demonstrates

that the hydrophobic gate is devoid of water.

As the electric field is increased to $\pm 100\,\mathrm{mV\,nm^{-1}}$, the channel exhibits frequent liquid-vapour oscillations (Figure 6.2A) and the time-averaged density in the hydrophobic gate takes on an intermediate value between a fully de-wetted and fully hydrated channel. Interestingly, a field of $-100\,\mathrm{mV\,nm^{-1}}$ appears to have a stronger effect than a field of $100\,\mathrm{mV\,nm^{-1}}$, suggesting that Beckstein openness depends on the direction of the electric field as well as on its magnitude. This idea will be pursued further below. At even higher values of the electric field of $\pm 200\,\mathrm{mV\,nm^{-1}}$, the channel pore exists nearly exclusively in the liquid state (Figure 6.2A) and the average water density lies near or above the density of bulk water throughout the channel (Figure 6.2B).

An external electric field can thus be used to control the phase behaviour of nanoconfined water, by influencing the relative probability of the liquid and vapour states. Figure 6.2C shows that the probability distribution of minimal water density is bimodal, with one peak near zero and a second peak near the density of bulk water. As can be seen, intermediate density values are observed only very rarely, which legitimises the theoretical treatment of the system in the context of a two-state model. Furthermore, this justifies the use of a universal cutoff of $n_{\mathrm{thres}} = 7.6\,\mathrm{nm^{-3}}$ in Equation 6.1 across all water models. The same cutoff also applies to the open conformation of the 5-$HT_3$ receptor (PDB ID: 6DG8, data not shown).

### Beckstein Openness Response

In Figure 6.3 the response of Beckstein openness to an external electric field is compared between the closed (PDB ID: 4PIR) and open (PDB ID: 6DG8) states of the 5-$HT_3$ receptor and across four different water models. For the closed state structure, this curve resembles a sigmoid, with Beckstein openness starting out near zero in the absence of an electric field and increasing to one at fields of large magnitude. The transitional regime of intermediate Beckstein openness has a width of $\sim 100\,\mathrm{mV\,nm^{-1}}$ for all four

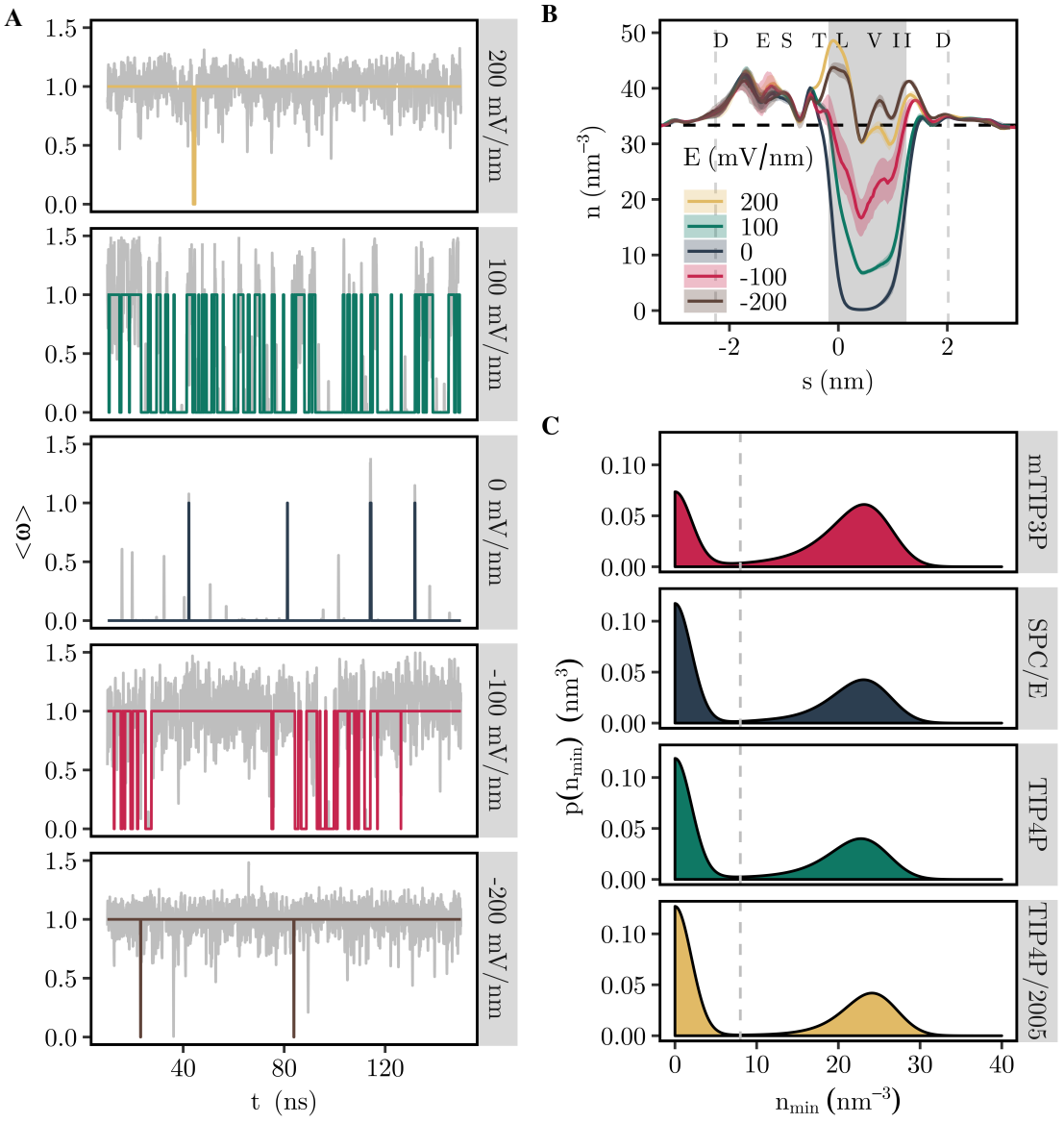



**Figure 6.2:** Electrowetting of the hydrophobic gate in the 5-HT$_3$ receptor. (A) Time series of Beckstein openness at five different electric field strengths. The grey line in the background represents the minimum water density in the hydrophobic gate normalised to the density of bulk water. (B) Time-averaged density profile in dependence of electric field strength. Confidence bands represent the standard deviation over three independent repeats. Shaded background and vertical dashed lines indicate the extent of the hydrophobic gate and transmembrane domain respectively. The vertical dashed line represents the density of bulk water. (C) Kernel density estimate of water density in the hydrophobic gate region aggregated over time and electric field strength. The dashed grey line indicates the threshold density for classifying the state of the channel as closed ($\omega(t) = 0$) or open ($\omega(t) = 1$). Data shown is based on simulations of the M2 helix bundle of the closed conformation of the 5-HT$_3$ receptor (PDB ID: 4PIR) using the mTIP3P water model unless otherwise noted.

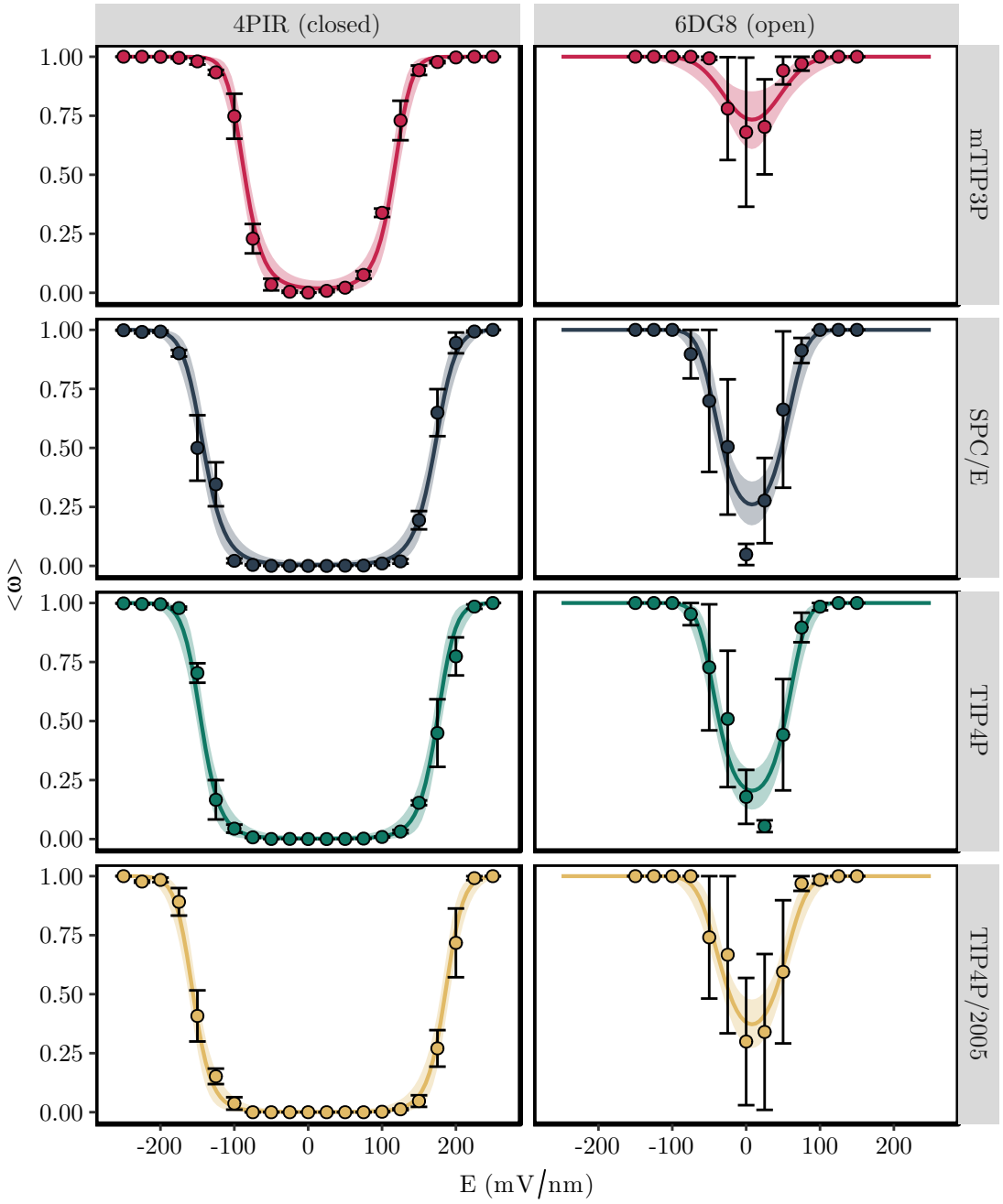



**Figure 6.3:** Beckstein openness, $\langle\omega\rangle$, in dependence of external electric field, $E$. Discrete data points and error bars represent the mean Beckstein openness and its standard error over three independent simulations respectively. Solid lines are the result of a Bayesian nonlinear multilevel model fitted to the data. The shaded area indicates the 95 % confidence interval of the fit. Simulations were performed on the M2 helix bundle of the closed (PDB ID: 4PIR) and open (PDB ID: 6DG8) conformations of the 5-$HT_3$ receptor and employed four different water models.

water models. However, the threshold field strength beyond which Beckstein openness exhibits an appreciable increase varies markedly across water models. For example, with the mTIP3P water model, pore wetting can be observed at a $\sim$50 mV nm$^{-1}$ weaker field than with any of the other models.

In case of the open state structure (PDB ID: 6DG8), the response of Beckstein openness still resembles a sigmoid, although Beckstein openness does not start out at zero. This is a consequence of the fact that for this conformation of the channel, the pore may be partially hydrated even in the absence of an electric field. Furthermore, the variation of Beckstein openness across repeats is significantly larger for this structure. This is likely due to the shorter simulation time of only 50 ns for this structure compared to the 150 ns carried out for the closed conformation of the channel.

It is important to point out that in the transitional regime the rate of increase in Beckstein openness is similar for both positive and negative electric fields, irrespective of which water model was used. However, the overall curve is shifted towards positive field strengths, which makes it asymmetric with respect to the $y$-axis. This can be seen particularly clearly for the closed state structure (PDB ID: 4PIR), where the threshold voltage for the onset of pore hydration is $\sim$25 mV nm$^{-1}$ higher for positive fields than for negative fields. Due to a smaller number of data points in the de-wetted regime, this effect is not as clear for the open state structure (PDB ID: 6DG8). However, the minimum of Beckstein openness appears to be shifted to slightly positive fields in this case as well.

This horizontal offset of the response curve is what motivated the introduction of the shifting parameter, $E_{\text{int}}$, into the thermodynamic model of Equation 6.5. Such a parameter is not present in the original model as introduced in Section 1.2.4, since Dzubiella and Hansen [48] considered only symmetric nanopores. Since $E_{\text{int}}$ must carry the same units as $E$, it can be tentatively associated with the intrinsic electric field caused by the charged amino acid residues of the 5-HT$_3$ receptor. This interpretation

will be scrutinised in Section 6.2.2.

### Bayesian Nonlinear Multilevel Model

The result of a Bayesian fit of the nonlinear multilevel model introduced in Equation 6.5 is included in Figure 6.3. Recall that $\Delta\Omega_0$ represents the free energy difference between the liquid and vapour states in the absence of an electric field, while $m$ denotes the strength of the coupling between time-averaged Beckstein openness and the magnitude of the field. $E_{\text{int}}$ accounts for the horizontal offset of the response curve. As can be seen, the model is generally in good agreement with the data, especially for the closed conformation of the 5-$HT_3$ receptor (PDB ID: 4PIR). For the open state structure (PDB ID: 6DG8), the model still captures the overall trend well, but appears to overestimate Beckstein openness at low electric field strengths. For the SPC/E and TIP4P models in particular, there is at least one data point (at field strengths of $0\,\text{mV}\,\text{nm}^{-1}$ and $25\,\text{mV}\,\text{nm}^{-1}$ respectively), which is not captured by the model.

One reason for this may be the larger statistical error caused by shorter simulation times for the open state structure as mentioned above. In addition, there are fewer data points at low Beckstein openness, because the open conformation of the pore is already partially hydrated in the absence of an electric field. During model fitting, this implicitly puts a stronger weight on the liquid state of the pore. Additional simulations in the low field strength regime could potentially remedy this issue.

The posterior probability distributions of all model parameters are displayed in Figure 6.4 and their values are also listed in Table 6.1. Importantly, this now allows a quantitative comparison of the zero-field free energy difference, $\Delta\Omega_0$, between the two conformational states of the channel. This was not possible in Chapter 5, because 150 ns long equilibrium simulations did not provide sufficient sampling to reliably estimate this parameter for the closed state structure (PDB ID: 4PIR).

As can been seen in Figure 6.4A, the zero-field free energy difference is close to zero

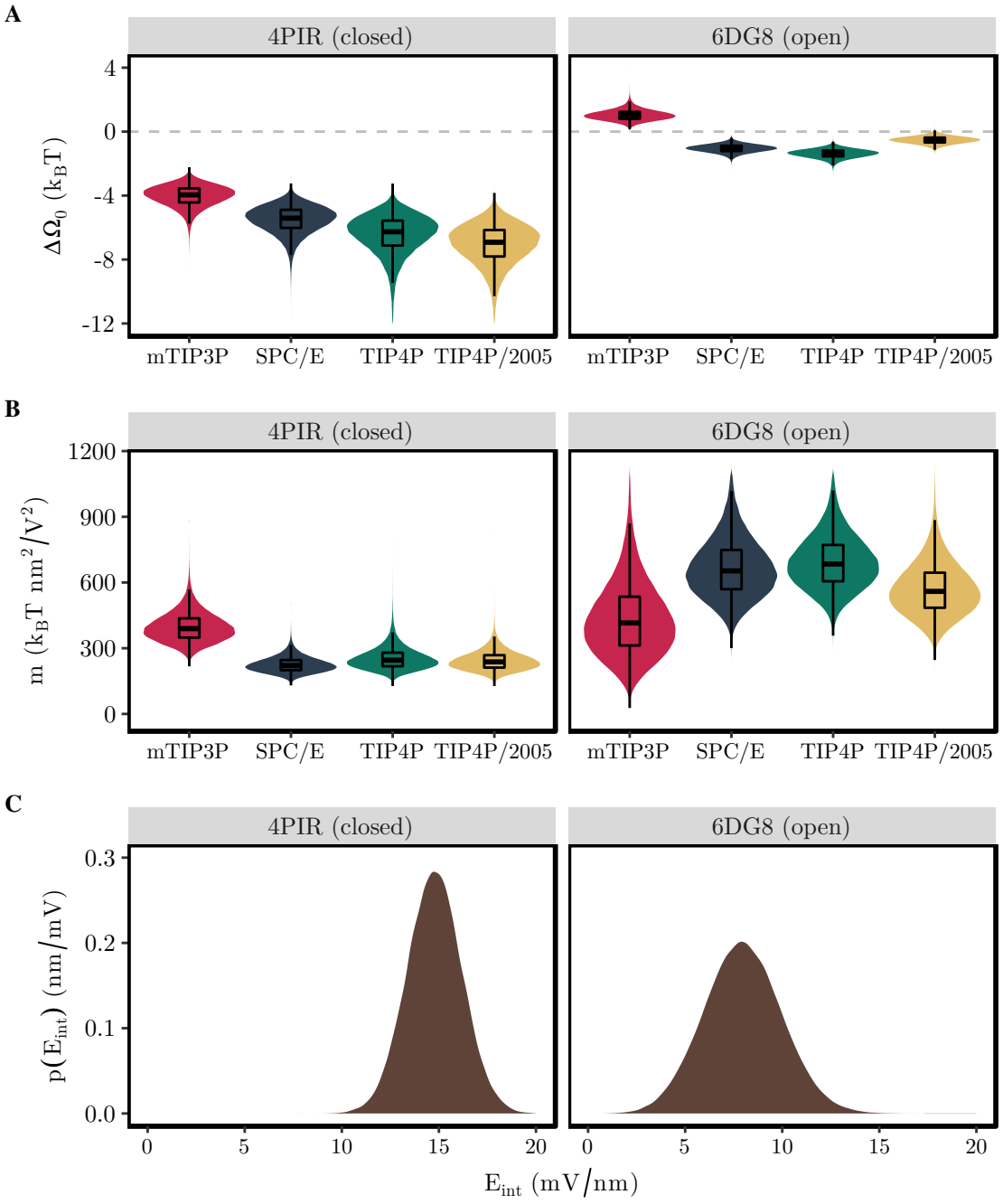



**Figure 6.4:** Posterior probability of Bayesian model parameters. (A) Posterior distribution of zero-field free energy difference, $\Delta\Omega_0$. Data are shown as box plot superimposed on a violin plot for each combination of water model and conformational state. The dashed line indicates equal probability of the liquid and vapour states. (B) Posterior distribution of electric field coupling parameter, $m$, for each combination of water model and conformational state. (C) Posterior distribution of intrinsic electric field, $E_{\text{int}}$. Since this parameter is independent of the water model (Section 6.1.3), only one kernel density estimate of the distribution is shown for each conformational state of the 5-$HT_3$ receptor.

| PDB ID | Water Model | $\Delta\Omega_0$ | $m$ | $E_{\text{int}}$ |
|---|---|---|---|---|
| 4PIR | mTIP3P | $-4.03 \pm 0.68$ | $396 \pm 67$ | — |
| 4PIR | SPC/E | $-5.51 \pm 0.88$ | $225 \pm 36$ | — |
| 4PIR | TIP4P | $-6.47 \pm 1.32$ | $253 \pm 52$ | — |
| 4PIR | TIP4P/2005 | $-7.08 \pm 1.31$ | $243 \pm 45$ | — |
| 4PIR | all | — | — | $14.80 \pm 1.41$ |
| 6DG8 | mTIP3P | $1.03 \pm 0.34$ | $436 \pm 171$ | — |
| 6DG8 | SPC/E | $-1.06 \pm 0.25$ | $668 \pm 139$ | — |
| 6DG8 | TIP4P | $-1.37 \pm 0.28$ | $695 \pm 126$ | — |
| 6DG8 | TIP4P/2005 | $-0.53 \pm 0.23$ | $571 \pm 122$ | — |
| 6DG8 | all | — | — | $7.97 \pm 1.99$ |

**Table 6.1:** Parameters of Bayesian nonlinear hierarchical model for Beckstein openness. Values listed represent the mean and standard deviation of the posterior distributions shown in Figure 6.4. Zero-field free energy difference, $\Delta\Omega_0$, electric field coupling, $m$, and intrinsic electric field, $E_{\text{int}}$, are given in units of $k_{\text{B}}T$, $k_{\text{B}}T\,\text{nm}^2\,\text{V}^{-2}$, and $\text{mV}\,\text{nm}^{-1}$ respectively.

across all water models for the open state structure (PDB ID: 6DG8). The mTIP3P water model slightly favours the liquid state by approximately $1\,k_{\text{B}}T$, while the other three models have a roughly $-1\,k_{\text{B}}T$ preference for the vapour state. By contrast, all water models strongly favour the vapour state in the closed conformation of the receptor (PDB ID: 4PIR), with zero-field free energy differences ranging from approximately $-4\,k_{\text{B}}T$ to about $-6\,k_{\text{B}}T$. Notably, it is the mTIP3P model that is closest to the liquid state in each case, in good agreement with previous results (Figure 5.7). However, the relative pore hydration preference of the other water models agrees less well across structures.

Figure 6.4B shows that the strength of the coupling between the hydration state of the pore and the electric field is also dependent on the structural conformation of the receptor. With the exception of the mTIP3P water model, this parameter is more than twice as large for the open state (PDB ID: 6DG8) than it is for the closed state (PDB ID: 4PIR). This difference can likely be attributed to the increased pore radius of the

open conformation, which corresponds to a larger volume entering into Equation 1.5. There appears to be no clear trend for how the coupling strength varies across water models, although it should be noted that this parameter is sensitive to intermediate values of Beckstein openness ($\langle\omega\rangle \approx 0.5$) and is therefore more likely to be affected by the comparatively larger error bars in the transitional regime (Figure 6.3).

Finally, the horizontal offset of the response of Beckstein openness to the electric field is quantified in Figure 6.4C. In the closed conformation (PDB ID: 4PIR) of the 5-$HT_3$ receptor, the value of the intrinsic field parameter is $\sim(14.80 \pm 1.41)\,\mathrm{mV\,nm^{-1}}$. This reduces to $\sim(7.97 \pm 1.99)\,\mathrm{mV\,nm^{-1}}$ in the open conformation (PDB ID: 6DG8). Due to the sign convention used in Equation 6.5, this field points from the extracellular to the intracellular domain. It therefore adds to the effective magnitude of an inward pointing external field ($E < 0$), while effectively reducing the impact of an outward pointing external field ($E > 0$). The following section will examine the origin of this intrinsic field in greater detail.

### 6.2.2 Intrinsic Electric Field of the 5-$HT_3$ Receptor

In the previous section, it was demonstrated that the response of time-averaged Beckstein openness to an external electric field is asymmetric with respect to the origin. This led to the introduction of an additional shifting parameter, $E_{\mathrm{int}}$, into the thermodynamic model fit to these data (Equation 6.5), which was associated with the electric field caused by the charged amino acids of the channel protein. While the interpretation of this parameter as an intrinsic electric field is sensible from the perspective of dimensional analysis, it is important to validate this assumption.

#### Electrostatic Potential

To this end, the electrostatic potential of the closed conformation 5-$HT_3$ receptor (PDB ID: 4PIR) was determined by numerically solving the Poisson-Boltzmann equation and

is shown in Figure 6.5. The profile of electrostatic potential along the centre line of the channel (Figure 6.5A) is characterised by two wells near the intra- and extracellular openings of the M2 helix bundle (positions C and A respectively), separated by a local maximum in the hydrophobic gate region (position B). Outside the channel pore, the electrostatic potential approaches zero, as would be expected for an electrolyte solution.

The distribution of charged amino acids shown in Figure 6.5B can qualitatively explain these features. The intracellular well in the vicinity of the pore-facing E$-1'$ residues is deeper than the extracellular well near the D20$'$ residues, because the pore is narrower at this point. In between, there are no charged, pore-facing residues so that the electrostatic environment is less favourable for cations. At the same time, the charges located on the D$-4'$, R0$'$, and K4$'$ residues are not directly exposed to the pore lumen and therefore have no significant influence on the electrostatic environment inside the channel.

As shown in Figure 6.5C, lipids change the electrostatic potential outside the protein, but have only a small influence on the electrostatic environment inside the pore. Consequently, the presence or absence of the lipid bilayer during the Poisson-Boltzmann calculations has only a minor influence on the shape of the electrostatic potential profile (Figure 6.5A), which is slightly lower near the extracellular opening, if lipids are included. This suggests that the presence of negatively charged residues pointing away from the pore, such as R0$'$ and K4$'$, can effectively shield the pore from the electrostatic influence of the lipid bilayer. Note that this agrees well with the results from Chapter 5, where it was shown that the composition of the lipid bilayer has a negligible influence on the induced dipole moment of water molecules inside the channel pore.

How does the electrostatic potential profile shown in Figure 6.5A relate to the intrinsic electric field postulated in the previous section? On the one hand, the local maximum of the potential inside the hydrophobic gate (position B) suggests that amino acid charges create no significant electric field perpendicular to the membrane plane in this region of the channel pore. On the other hand, the overall shape of the profile indicates that

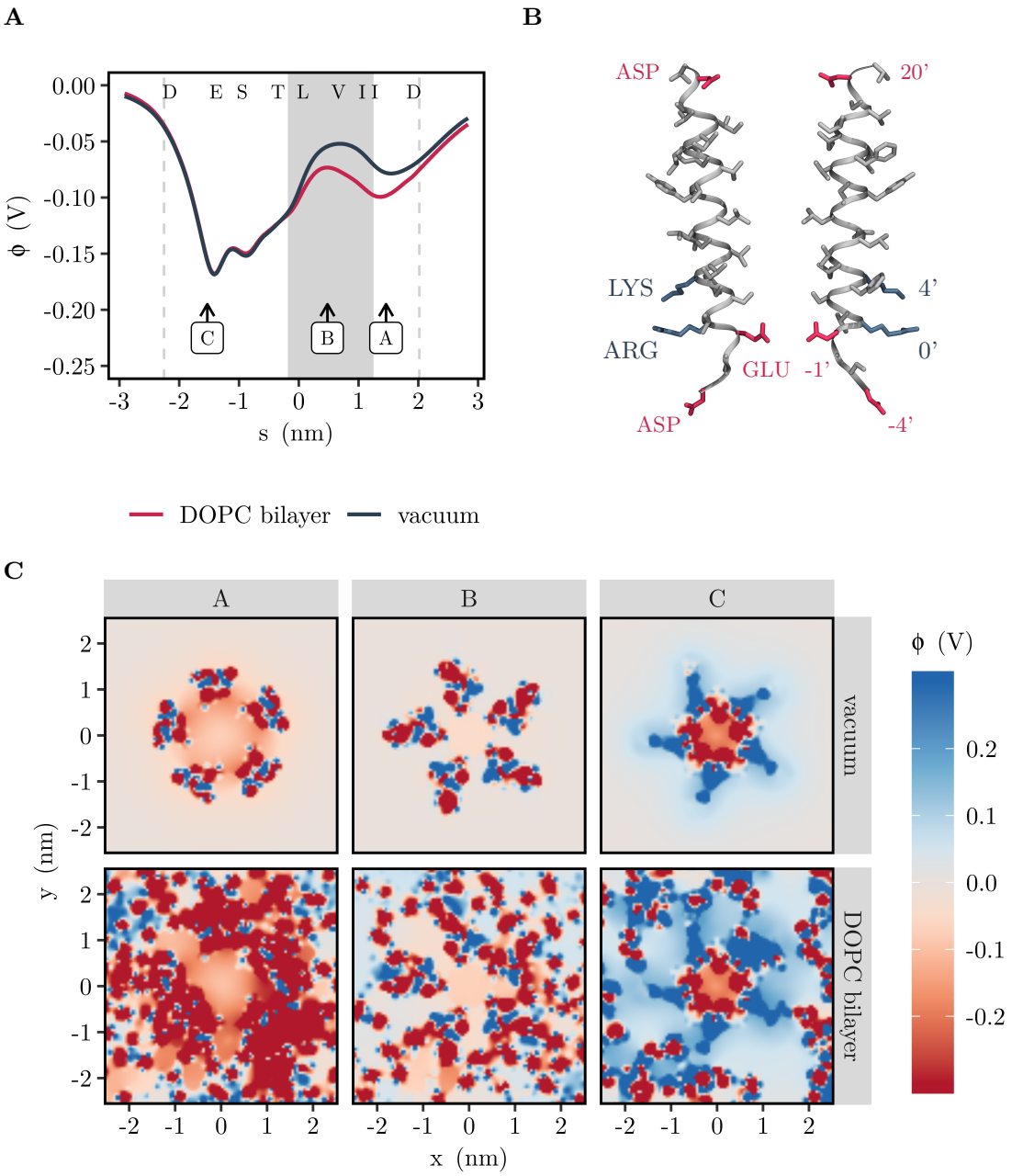



**Figure 6.5:** Electrostatic potential of the M2 helix bundle of the closed state 5-$HT_3$ receptor (PDB ID: 4PIR). (A) Profile of electrostatic potential along the pore centre line, calculated for a helix bundle in vacuum and in a DOPC bilayer. (B) Structure of the M2 helix bundle. Positively and negatively charged amino acids are shown in blue and red licorice representation respectively, all other residues are coloured grey. For visual clarity, only two subunits are shown. (C) Heat map of electrostatic potential for three lateral cross sections of the helix bundle. The position of these cross sections is indicated by labelled arrows in panel A.

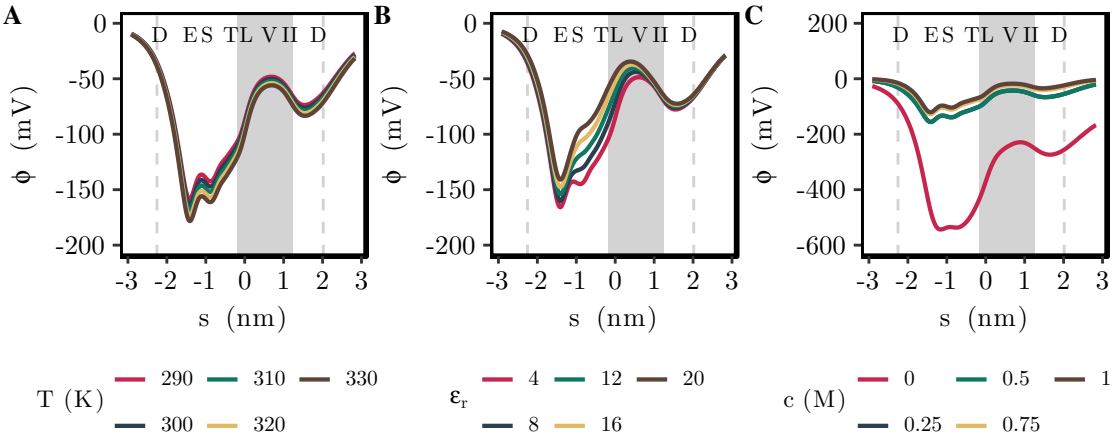



**Figure 6.6:** Parameter dependence of the electrostatic potential profile. (A) Influence of temperature, $T$. (B) Influence of the relative permittivity of the protein, $\varepsilon_\mathrm{r}$. (C) Influence of ion concentration, $c$. All calculations were carried out on the M2 helix bundle of the closed state 5-$HT_3$ receptor (PDB ID: 4PIR) in the absence of lipids.

there is a substantial potential gradient between the intra- and extracellular openings of the M2 helix bundle. Since it is not clear that a continuum electrostatics calculation can provide quantitatively accurate results for the potential inside a molecular scale protein cavity, the changes of the potential over larger scales may be more reliable than its small scale variations.

In fact, the electrostatic potential increases from approximately $-160\,\mathrm{mV}$ in the intracellular well (position C) to nearly $-75\,\mathrm{mV}$ in the extracellular well (position A) over a distance of $\sim 3\,\mathrm{nm}$. This corresponds to an electric field of $\sim 28\,\mathrm{mV\,nm^{-1}}$ (or $\sim 20\,\mathrm{mV\,nm^{-1}}$ when the DOPC bilayer is included), which is reasonably close to the value of $E_\mathrm{int} = (14.80 \pm 1.41)\,\mathrm{mV\,nm^{-1}}$ found from the nonlinear fit shown in Figure 6.4C. The continuum electrostatics calculations thus strengthen the view that an intrinsic electric field causes the asymmetry in the electric field response of Beckstein openness.

Figure 6.6 addresses the question of how sensitive these results are to the parameter choices underlying the Poisson-Boltzmann calculations. As can be seen, temperature has

only a negligible effect on the electrostatic potential profile (Figure 6.6A). Similarly, the relative permittivity of the protein only influences the electrostatic potential between the intracellular potential well ($s \approx -1\,\mathrm{nm}$) and the hydrophobic gate ($s \approx 0\,\mathrm{nm}$), where a higher permittivity increases the potential (Figure 6.6B). However, the overall shape of the profile is not impacted by the relative permittivity.

By contrast, the ionic strength of the electrolyte solution has a similarly small influence on the electrostatic potential profile only if it is not reduced to zero (Figure 6.6C). In the absence of mobile charges (i.e. $c = 0$), the electrostatic potential is lowered significantly throughout the entire pore and the intracellular well reaches a depth of approximately $-500\,\mathrm{mV}$. However, such a scenario is unphysiological and the strong dependence of the potential on the concentration of ions is primarily an indication of the limited applicability of continuum electrostatics at the molecular scale.

### Side Chain Reorientation

In principle, the asymmetry of the openness response curve (Figure 6.3) could also be due to a reorientation of amino acid side chains under the influence of the external electric field. Side chain rotations could change both the radius of the pore and the effective hydrophobicity of its surface, which would in turn impact Beckstein openness. If the reorientation probability was biased by the direction of the electric field, for example because steric clashes prevent the flipping of side chains in one direction, this bias could lead to an asymmetric threshold field strength for pore hydration.

Figure 6.7 displays how the preferential orientation of the three residues lining the hydrophobic gate varies under the influence of the electric field. As can be seen, there is no significant dependence of the preferential side chain orientation on the electric field. Any observable trend is smaller than the statistical variation at a given magnitude of the field. This behaviour is reasonable, as these residues carry no net charge and, being hydrophobic, have only a small net dipole moment, so that their interaction

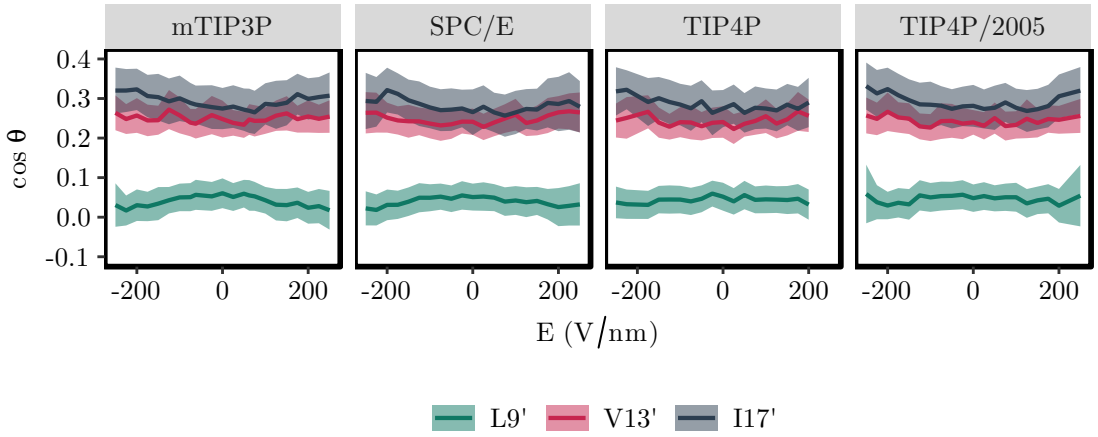



**Figure 6.7:** Orientation of amino acid side chains facing the hydrophobic gate of the closed state 5-$HT_3$ receptor (PDB ID: 4PIR) in dependence of external electric field strength. Confidence bands represent the orientational variation over time aggregated over three independent simulations at each field strength. While the orientation of all residues varies over time, there is no systematic variation with field strength. Side chain orientational variation is also independent of the water model.

with the electric field is necessarily weak. Consequently, electric-field-biased side chain reorientation is unlikely to play an important role here.

### 6.2.3 Ionic Currents

Having established that a sufficiently strong electric field can hydrate a hydrophobic gate, it is interesting to consider in how far the hydration state of the pore influences ion conduction through the channel. Figure 6.8 uses the example of the closed conformation of the 5-$HT_3$ receptor (PDB ID: 4PIR) in combination with the mTIP3P water model to illustrate how the total amount of charge transferred through the channel increases over time. This is shown for five different magnitudes of the applied electric field.

As can be seen, virtually no charge is transferred at field strengths $<150\,\mathrm{mV\,nm^{-1}}$, but if the electric field is increased above this threshold, charge is steadily flowing through the channel. During each simulation, the total amount of charge moved through the channel increases approximately linearly over time. This indicates that the ionic current is nearly

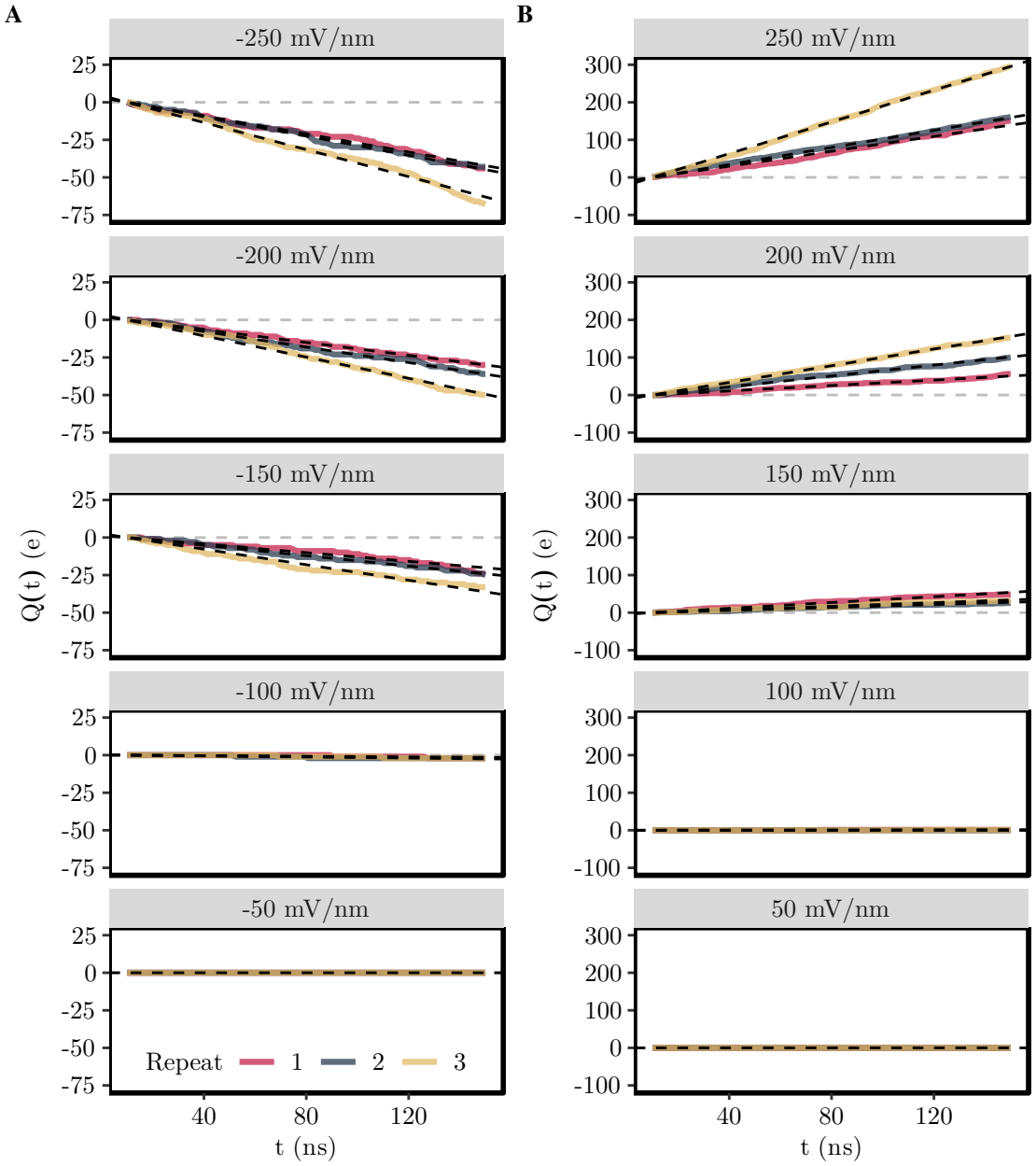



**Figure 6.8:** Cumulative charge flow through the M2 helix bundle of the closed state 5-$HT_3$ receptor (PDB ID: 4PIR) due to $Na^+$ ions. Coloured lines indicate the charge flow measured during three independent 150 ns long MD simulations employing the mTIP3P water model, while dashed black lines represent a linear fit to these data. The dashed grey lines serve as reference and indicate the case of no electric current. (A) Charge flow under the influence of a negative electric field ($E < 0$). Positive charges are driven from the extracellular space to the intracellular domain. (B) Charge flow under the influence of a positive electric field ($E > 0$). The direction of charge flow is reversed here.

constant and it can therefore be determined through a linear fit. Notably, the rate of ion conduction may vary across repeats and the degree of variation appears to increase with increasing field strength. This is a consequence of the fact that for a Poisson process like the counting of conduction events carried out here the standard deviation increases with the mean. An overall similar relationship between charge transfer and field strength was observed for the closed conformation of the 5-$HT_3$ receptor, as well as for all other water models.

Figure 6.9 depicts how the average current, decomposed into the contributions of $Na^+$ and $Cl^-$ ions, varies with the external electric field. Intriguingly, all curves share some common features. Firstly, the current due to either ion species is approximately zero for low electric fields and begins to increase in magnitude only when the field exceeds the critical value required to hydrate the hydrophobic gate (Figure 6.3). Following an initial nonlinearity, the current increases linearly with field strength for large electric fields. This implies that once the pore is hydrated, it effectively acts as an Ohmic resistance, as is commonly assumed in theoretical models of ion channel conductance.[2] These results demonstrate directly that pore de-wetting prevents the formation of an ionic current. Importantly, it can be ruled out that the onset of ionic current merely coincides with the onset of pore hydration or that the permeation of ions itself causes pore hydration, since the threshold field strength for hydration differs by water model.

Secondly, the current due to $Na^+$ ions is significantly larger than the current due to $Cl^-$ ions, in agreement with the experimentally well-established cation selectivity of the 5-$HT_3$ receptor. This is likely due to the negatively charged residues near the intracellular (D$-4'$, E$-1'$) and extracellular (D$20'$) openings of the M2 helix bundle (Figure 6.5B), which make an entry into the pore highly unfavourable for $Cl^-$ ions, but highly attractive for $Na^+$ ions. Note that it was shown experimentally by Thompson and Lummis [339] that mutating E$-1'$ into a neutral residue removes the cation selectivity of the 5-$HT_3$ receptor.

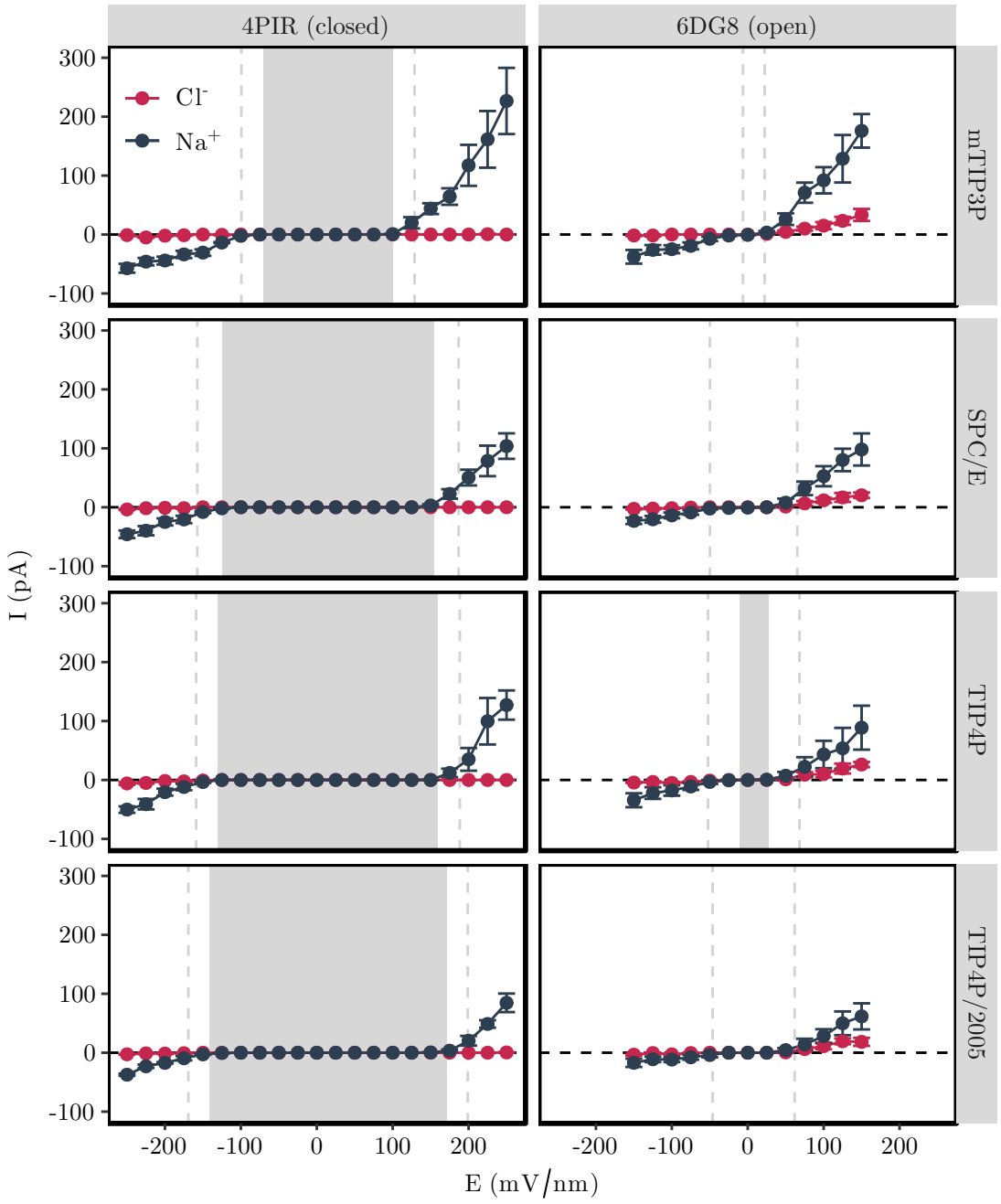



**Figure 6.9:** Ionic current through the 5-$HT_3$ receptor M2 helix bundle in dependence of the external electric field. Background shading indicates the regime where the hydrophobic gate is mostly dewetted ($\langle\omega\rangle < 0.25$), while the vertical dashed lines indicate the transition to a mostly hydrated ($\langle\omega\rangle > 0.75$) pore. These thresholds are calculated from the Bayesian multilevel model shown in Figure 6.3 and background shading is omitted in cases where the fit curve does not fall below 0.25.

Thirdly, the channel appears to act as a rectifier: the ionic current for a positive electric field ($E > 0$) is significantly larger than the current at a negative field ($E < 0$) of comparable magnitude, provided that the field is sufficiently strong to hydrate the channel pore. This is true despite the fact that an overall stronger positive field is required to hydrate the pore and is therefore unlikely to be a consequence of the hydrophobic gate. One possible explanation for this behaviour is the larger number of negatively charged residues near the intracellular opening of the M2 helix bundle (Figure 6.5B). This could cause a higher density of $Na^+$ ions near the intracellular opening, thus increasing the reservoir of ions ready to be conducted under the influence of the external field. However, it should be noted that this rectifying behaviour is physiologically irrelevant, as it occurs only at supraphysiological field strengths.

Finally, it should be noted that the open conformation of the 5-$HT_3$ receptor (PDB ID: 6DG8) exhibits a larger conductance than the closed conformation (PDB ID: 4PIR). This manifests itself as a larger ratio of current to field strength, but only in the large-field regime, where the hydrophobic gate is hydrated. It should also be noted that a direct comparison of these currents to electrophysiological data is not sensible, since the structural model used in these simulations comprises only the M2 helix bundle, while the extra- and intracellular domains will likely influence the absolute magnitude of the observed current.

### 6.2.4 Kinetics of Liquid-Vapour Transitions

In the previous sections it was shown that an external electric field influences the hydration probability of a channel pore by changing the free energy difference between its liquid and vapour states. This section examines how the free energy difference is controlled by the underlying rate constants that determine the kinetics of liquid-vapour oscillations inside a hydrophobic gate. Figure 6.10 shows how the mean survival times of the liquid state, $\tau_\mathrm{l}$, and vapour state, $\tau_\mathrm{v}$, vary with with the free energy difference

between these two states. Noting the logarithmic scaling of the $y$-axis, it can be seen that $\tau_\mathrm{l}$ grows exponentially with $\Delta\Omega$, while $\tau_\mathrm{v}$ decreases exponentially at approximately the same rate. This trend appears to be very similar for both conformational states of the channel as well as across all four water models.

It should also be noted that there is a significantly higher degree of variation around the exponential trend in the case of the open conformation (PDB ID: 6DG8). This is due primarily to the shorter simulation time of only 50 ns compared to the 150 ns carried out for the closed conformation (PDB ID: 4PIR). Furthermore, there are fewer data points for the open state structure. This is caused by the fact that the channel pore does not fully de-wet even in the absence of an electric field, so that fewer simulations fall in the intermediate hydration range ($\Delta\Omega \approx 0$), in which survival times can be estimated reliably.

Both curves in Figure 6.10 necessarily cross one another at $\Delta\Omega = 0$, as an equal probability for both states implies equal survival times. The crossover point thus defines a characteristic time scale for liquid-vapour oscillations, which for the 5-$HT_3$ receptor takes on a value of $\tau_\mathrm{lv} = \tau_\mathrm{l}(\Delta\Omega = 0) = \tau_\mathrm{v}(\Delta\Omega = 0) \approx 5\,\mathrm{ns}$. Importantly, this value is nearly identical for both conformational states of the receptor, despite their significantly different pore radii. Furthermore, $\Delta\Omega = 0$ occurs at very different magnitudes of the electric field. This indicates that a time scale of $\sim$5 ns is likely to characterise liquid-vapour oscillations of all pores that share the characteristic dimensions and surface hydrophobicity of the 5-$HT_3$ receptor, i.e. most other ion channels. Moreover, there is some slight variation of the oscillation time scale across water models, but $\tau_\mathrm{lv}$ always falls in the range between 1 ns and 10 ns, indicating a limited force field sensitivity.

### Minimum Simulation Time

Combined with the observation that $\tau_\mathrm{l}$ and $\tau_\mathrm{v}$ exhibit an inverse dependency on $\Delta\Omega$, knowledge of the characteristic oscillation time scale provides a way to estimate how

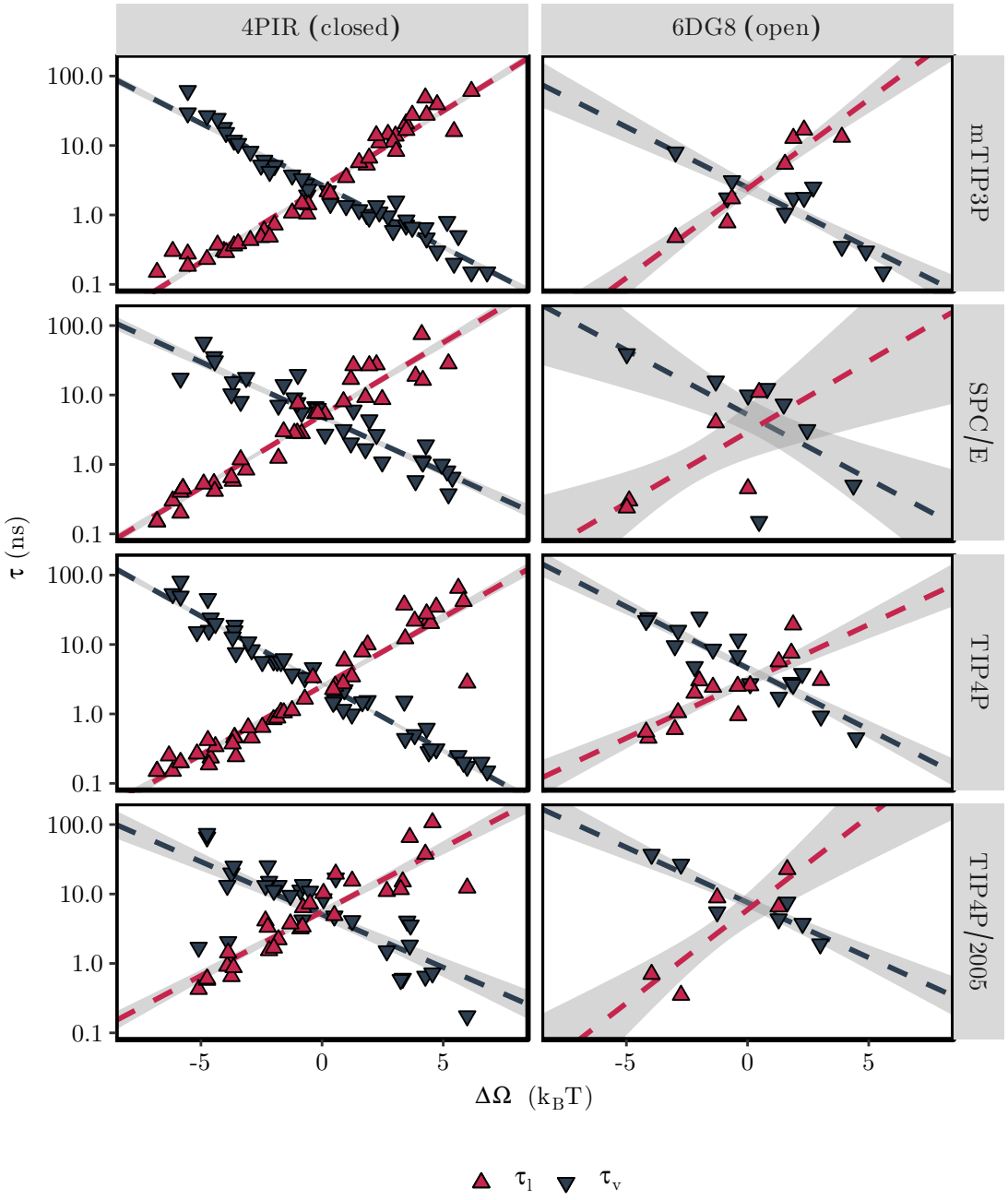



**Figure 6.10:** Kinetics of liquid-vapour transitions under the influence of an external electric field. The dependence of the mean survival times of the liquid ($\tau_{\text{liquid}}$) and vapour ($\tau_{\text{vapour}}$) states on the free energy difference ($\Delta\Omega$) between these states is shown. Each data point corresponds to an individual simulation, while the dashed line represents a linear fit to all data points, with confidence bands indicated as shaded background. The free energy difference was estimated from the mean Beckstein openness, while the survival times were determined from the time series of Beckstein openness. Simulations with fewer than two observed transitions are not included in this plot.

long an equilibrium MD simulation needs to be in order to classify a hydrophobic gate as closed or open. If the free energy difference is small ($\Delta\Omega \approx 0$), many liquid-vapour oscillations will be observable even during a short simulation, making it possible to directly estimate $\Delta\Omega$ as well as $\tau_\mathrm{l}$ and $\tau_\mathrm{v}$. In practice, a total simulation time that is an order of magnitude larger than the characteristic time scale, i.e. $10\,\tau_\mathrm{lv} \approx 50\,\mathrm{ns}$, can be considered sufficient.

However, if the free energy difference is large ($\Delta\Omega \gtrsim \pm 5\,k_\mathrm{B}T$), liquid-vapour transitions cannot be observed during the course of such a short simulation (Figure 6.10). If in this case the pore is seen to exist in the vapour state for time $T_\mathrm{v} = T_\mathrm{sim} \gg \tau_\mathrm{lv}$, the free energy difference must be negative ($\Delta\Omega \ll 0$). Conversely, if the pore is seen to exist in the liquid state for time $T_\mathrm{l} = T_\mathrm{sim} \gg \tau_\mathrm{lv}$, the free energy difference is necessarily positive ($\Delta\Omega \gg 0$). Note that this would not be the case if $\tau_\mathrm{l}$ and $\tau_\mathrm{v}$ were both positively correlated with $\Delta\Omega$.

Thus, while a short ($\sim$50 ns) simulation would not be enough to estimate the free energy difference, it would still be possible to establish whether it lies below $-5\,k_\mathrm{B}T$ or above $+5\,k_\mathrm{B}T$. As was shown in the foregoing sections, a free energy difference of $-4\,k_\mathrm{B}T$ is already sufficient to functionally close a pore to ions even under a strong electric field in excess of $100\,\mathrm{mV\,nm^{-1}}$ (Figures 6.4 and 6.9). Consequently, a short equilibrium MD simulation would suffice to identify a closed hydrophobic gate, even if liquid-vapour oscillations cannot be observed directly.

Note that the foregoing considerations provide an estimate for the minimum duration of an individual MD simulation required to capture the dynamics of liquid-vapour oscillations in ion channel pores. It is still necessary to carry out multiple independent simulation repeats in order to obtain statistically meaningful results. Therefore, the cumulative simulation time would be larger than the $\sim$50 ns proposed above.

### Ion Dwell Times

It is also worthwhile to compare the duration of the liquid and vapour states with the typical dwell time of ions inside the channel pore. Figure 6.11 shows the relationship between the strength of the electric field and ion dwell time. As can be seen, the average ion dwell time decreases with increasing magnitude of the electric field. This indicates that the increase in ionic current shown in Figure 6.9 is at least in part due to faster motion of individual ions through the channel pore. Furthermore, ions require less time to pass the open conformation of the 5-$HT_3$ receptor (PDB ID: 6DG8), which is in agreement with the higher current (at similar field strength) observed for this structure (Figure 6.9). In the case of the closed conformational state (PDB ID: 4PIR), ion dwell times seem to be much lower for positive fields than for negative fields (Figure 6.11). This might be related to the current rectification observed in Figure 6.9, although this effect is less pronounced for the open conformation (PDB ID: 6DG8).

Noting that the ion dwell time refers to the time an ion spends inside the entire pore, rather than in the hydrophobic gate region alone, the data shown in Figure 6.11 provides a valuable consistency check. For ions that pass the closed conformation (PDB ID: 4PIR) in its partially hydrated state ($0.25 < \langle\omega\rangle < 0.75$), the average dwell time is comparable to the typical time scale of liquid-vapour oscillations ($\tau_{\mathrm{lv}} \approx 5\,\mathrm{ns}$). This implies that intermittent periods during which the channel exists in the liquid state last sufficiently long to allow the passage of individual ions.

For the open conformation (PDB ID: 6DG8), the typical ion dwell time lies below $\sim$1 ns. This is less than the average life time of the liquid state, which exceeds $\sim$1 ns even for a slightly negative free energy difference ($\Delta\Omega > -2.5\,k_{\mathrm{B}}T$). Consequently, even though most water models predict a significant dehydration of the pore in the absence of an electric field (Figure 6.3), this conformation might still allow the occasional passage of ions, in agreement with its interpretation as a functionally open state. However, it should be noted that the observed current at physiological field strengths is very low

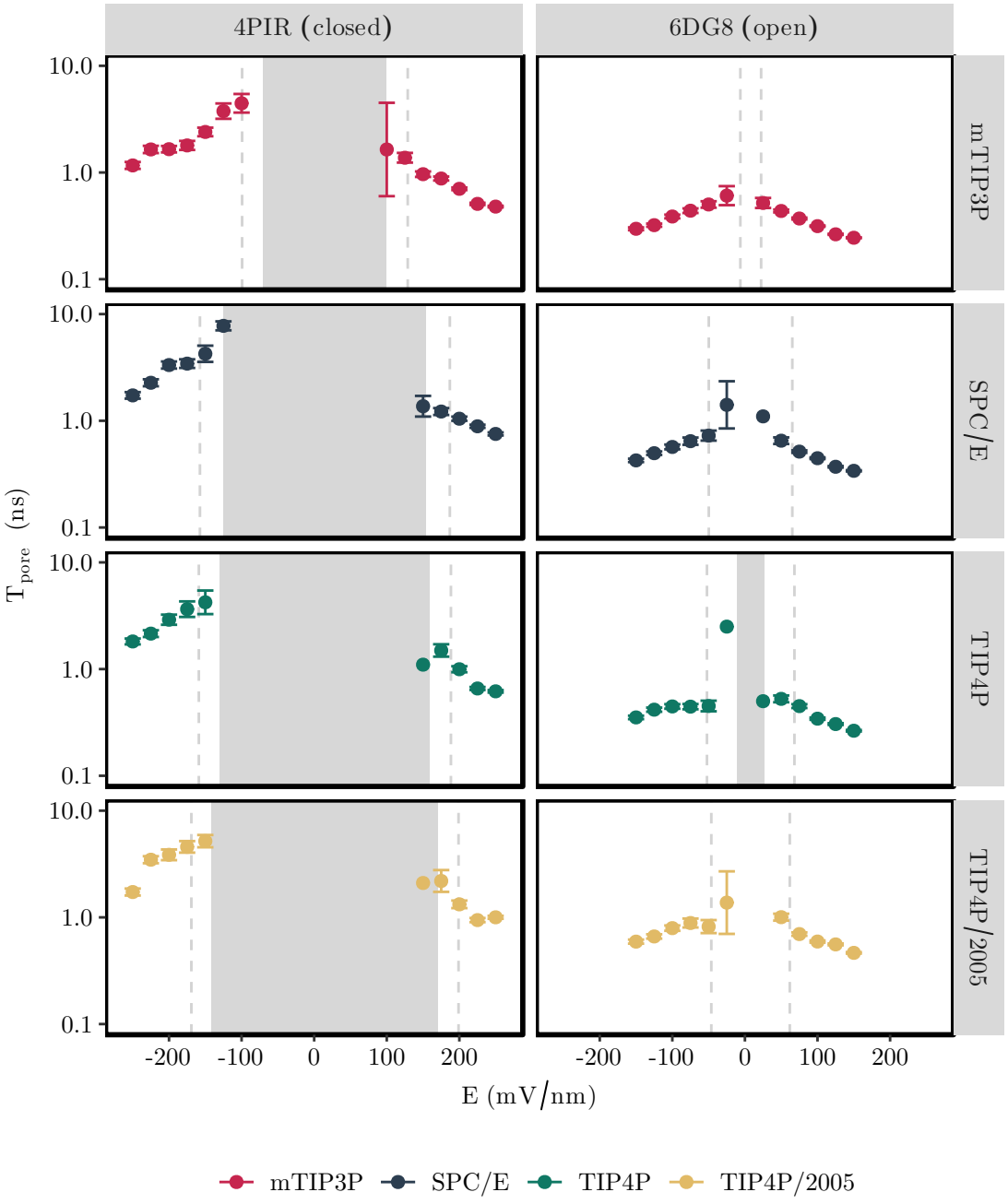



**Figure 6.11:** Typical dwell time of ions inside the pore of the 5-$HT_3$ receptor M2 helix bundle in dependence of external electric field strength. Background shading indicates the regime in which the channel is de-wetted ($\langle\omega\rangle < 0.25$), while the dashed lines mark the transition to a mostly hydrated channel ($\langle\omega\rangle > 0.75$). These thresholds are calculated from the Bayesian multilevel model shown in Figure 6.3 and background shading is omitted in cases where the fit curve does not fall below 0.25. Ions that exit the pore in the direction from which they entered are not included and no distinction is made between $Na^+$ and $Cl^-$ ions.

(Figure 6.9), which is likely due to the overall low conductivity of the 5-$HT_3$ receptor.

## 6.3 Discussion

In Section 6.2.1 it was shown that an external electric field representative of the influence of a transmembrane potential can induce wetting of the hydrophobic gate in the 5-$HT_3$ receptor. However, a field strength of $\sim$200 mV nm$^{-1}$ is required to fully hydrate the pore of the closed conformation of the channel. This corresponds to a potential difference of $\sim$1.7 V and thus significantly exceeds any transmembrane voltage an ion channel would experience under normal physiological conditions. Furthermore, the transitional regime between a completely de-wetted and a fully hydrated channel extends over $\sim$50 mV nm$^{-1}$, corresponding to a voltage of $\sim$425 mV. For the functional annotation of ion channels this implies that it is highly unlikely that an ion channel, which completely de-wets in an equilibrium simulation, will be falsely assigned to the closed state.

Section 6.2.1 also demonstrated that a simple thermodynamic model can quantitatively describe the increase in pore hydration probability with increasing electric field. However, to accurately account for the asymmetry of the Beckstein openness response curve (Figure 6.3), the model proposed by Dzubiella and Hansen [48] needed to be amended by an additional horizontal offset. The corresponding parameter, $E_{\mathrm{int}}$, was linked to the intrinsic electric field caused by the charged residues of the 5-$HT_3$ receptor. In Section 6.2.2, Poisson-Boltzmann electrostatics calculations were used to confirm the existence of a net electric field across the pore. Furthermore, it was ruled out that preferential reorientation of the amino acid side chains facing the hydrophobic gate contributes to the direction dependence of the wetting response.

It is important to point out that $E_{\mathrm{int}}$ represents a property of the receptor protein. Consequently, it does not account for a possible direct influence of the direction of the external electric field on the water molecules confined inside the channel pore. In the simulations of Daub et al. [319] as well as Bratko et al. [320], the direction of the electric

field influenced the orientation of interfacial water molecules and led to a distortion of their hydrogen bonding network. While it is possible that these effects are also relevant for water confined in the 5-$HT_3$ receptor, the fit shown in Figure 6.3 suggests that their magnitude is likely negligible. Firstly, the observed shift of the response curve (Figure 6.3) is well described by the current model and, secondly, there is no apparent difference in the magnitude of this shift across water models.

Moreover, Daub et al. [319] as well as Bratko et al. [320] considered planar geometries and observed a direction dependence when the electric field was varied between orientations parallel and perpendicular to the (planar) liquid-solid interface. By contrast, only two opposite orientations of the electric field were considered here, both of which are parallel to the approximately cylindrical pore surface. Note that these are the only biologically relevant field orientations, as an ionic imbalance across a lipid bilayer can only give rise to an electric field that is perpendicular to the membrane.

The finding that the threshold voltage required for hydrating a hydrophobic gate depends on the orientation of the electric field is an interesting perspective for the design of artificial nanopores. The calculations presented in Section 6.2.2 suggest that an excess of five elementary charges near one opening of the channel pore suffices to create a substantial intrinsic electric field acting in the transmembrane direction. By functionalising the openings of a hydrophobic pore with static charges, it might thus be possible to design a nanochannel that permits transport in only one direction, thus rectifying currents. Importantly, this rectifying property would not only apply to ions, but to water molecules as well. This extends the concept of the electric-field induced smart water gates proposed by Xiao et al. [324] and would permit the creation of membrane systems that are semi-permeable to water.

In Section 6.2.4, the electric field was used to analyse the kinetics of liquid-vapour oscillations by systematically biasing the channel pore towards specific points in phase space. It was found that the typical life time of the liquid state increases exponentially

with increasing free energy difference between the two hydration states, while simultaneously the survival time of the vapour state decreases exponentially. From this inverse relationship it was concluded that relatively short simulations of ∼50 ns duration, as used throughout this thesis, are sufficient to determine the functional status of an ion channel with a hydrophobic gate.

However, it is possible that the application of an external electric field influences the kinetic constants of the liquid-vapour transition in a way that corresponding changes in pore radius or surface hydrophobicity would not. For example, the electric field enforces a preferential orientation of water molecule dipoles and thus introduces a higher degree of order into the system. However, an equal probability of both hydration states ($\Delta\Omega = 0$) is reached at very different magnitudes of the electric field in the open and closed conformations of the 5-$HT_3$ receptor, while the associated time scale ($\tau_{lv}$) does not vary substantially. This suggests that an influence of the electric field on the typical life times of the system, beyond its contribution to the free energy difference ($\Delta\Omega$), is likely limited.

Furthermore, liquid-vapour oscillation kinetics were previously studied by Beckstein and Sansom [41] through simulations of model nanopores, in which the free energy difference was controlled by varying the pore radius. Here it was found that the life time of the liquid state increases exponentially with pore radius (and hence with the free energy difference), in agreement with the results presented here. By contrast, the survival time of the vapour state remained approximately constant for radii below ∼0.6 nm (the threshold below which de-wetting occurred) and then decreased significantly for larger radii.[41] Nonetheless, the main conclusion regarding the required minimum simulation time derived in Section 6.2.4 would still hold in this case. This is because the characteristic time scale ($\tau_{lv}$) would not be informative only if both survival times ($\tau_l$ and $\tau_v$) simultaneously increase or simultaneously decrease with the free energy difference.

## 6.4 Summary and Conclusions

This chapter investigated under which conditions a transmembrane potential can cause the hydration of a hydrophobic gate. It was shown that an external electric field can indeed hydrate the pore of the 5-$HT_3$ receptor, even in its closed conformational state. However, the field required for fully opening the hydrophobic gate corresponds to a transmembrane potential of $\sim$1.7 V and thus significantly exceeds the physiologically relevant voltage range. This shows that the vapour state of the closed hydrophobic gate in the 5-$HT_3$ receptor can be very stable and is not strongly influenced by external factors.

In addition, it was shown that ionic currents through the 5-$HT_3$ receptor pore are suppressed even at very strong electric fields. A significant number of conduction events was only observed at voltages above the electrowetting threshold, when the hydrophobic gate was hydrated. Notably, this threshold depends on the choice of water model, which indicates that the primary impediment to ion permeation is the hydration state of the pore rather than its structural features. Collectively, these observations demonstrate that electrowetting is unlikely to be physiologically relevant and that the functional annotation derived from equilibrium MD simulations is robust.

It was also found that the wetting response of the 5-$HT_3$ receptor pore depends on the direction of the external electric field. Through a combination of statistical modelling and continuum electrostatics calculations based on the Poisson-Boltzmann equation it was shown that the most likely cause for this asymmetry is the unequal distribution of charged protein residues along the ion channel pore. These amino acid charges create an additional electric field intrinsic to the channel protein that either stacks with the external field or counteracts it. In the case of the 5-$HT_3$ receptor, this leads to an opening of the hydrophobic gate at a lower voltage, if the applied field points from the extracellular space towards the intracellular domain.

Finally, the electrowetting effect has enabled a quantitative analysis of the reaction

kinetics of the phase transitions of nanoconfined water. As was shown in Section 6.2.4, liquid-vapour oscillations in ion channels occur on a characteristic time scale of $\sim$5 ns, when both phases are equally probable. In addition, the survival times of the liquid and vapour states have inverse relationships to the free energy difference between the two hydration states. Consequently, equilibrium simulations of a duration of several times the characteristic time scale, i.e. $\sim$50 ns, are sufficiently long to identify a hydrophobic gate in an ion channel structure. This result once more emphasises the high-throughput character of the functional annotation protocol introduced in Chapter 3.

I would have written a shorter thesis, but I did not have the time.

*(After Blaise Pascal)*

# 7

# Summary and Conclusions

Driven by advances in cryo-EM, research into the structural biology of ion channels has been fruitful over the past decade: since the beginning of this project in 2016, over 500 new channel structures have been deposited in the PDB. However, the hydrophobic gating effect, reported by a large number of computational studies[50,56,57,61,63,73–77] and supported by experimental evidence,[51,81–83,323,324] presents a challenge for interpreting this wealth of structural data. Hydrophobic gates prevent the flow of ions not by steric occlusion of the permeation pathway through an ion channel, but by the evaporation of water from the channel pore. Consequently, they cannot be identified by measurements of just the physical dimensions of a pore and thus cannot be detected by many of the tools conventionally applied to assign a functional status to an ion channel structure. The aim of this thesis was therefore to develop an efficient computational approach to the functional annotation of channel structures based on MD simulation. Its main results are summarised below.

## 7.1 The Channel Annotation Package

In Chapter 3, the Channel Annotation Package (CHAP) was introduced. In contrast to existing tools, CHAP calculates a pore hydrophobicity profile in addition to a pore

radius profile. If a channel contains a constriction lined by hydrophobic residues that may act as a hydrophobic gate, CHAP can also determine water density and free energy profiles from equilibrium MD simulations of that channel. Through this, CHAP permits the identification of hydrophobic gates that would otherwise remain undiscovered if only the pore dimensions were analysed.

CHAP has been made freely available to the structural biology community as an open-source software tool.[26] It is integrated with scripts for visualising results using the popular molecular graphics programmes VMD and PyMOL. Its source code can be obtained from Github[247] and a ready-to-use installation is available via the structural biology software consortium SBGrid.[250,251] Initial applications of this new tool include structural studies of eukaryotic[232] and bacterial[252] transport proteins as well as simulation studies of artificial nanopores.[254]

Further development of CHAP will focus on two areas: (i) the implementation of additional features of the `chap` command line utility and (ii) an improved accessibility of the tool for end users. While CHAP is already capable of identifying highly curved pores, it is currently limited to non-branching permeation pathways. The ability to identify branched pores could be implemented by replacing the current probe-based path finding algorithm with a procedure based on Voronoi tessellation. In addition, CHAP currently operates as a single-threaded programme and could in the future be extended to allow parallel processing of MD trajectories. Support for visualisation in additional common molecular graphics programmes, such as UCSF Chimera,[340] would also be a further sensible extension.

Accessibility could be improved through the development of a web-based interface for CHAP that allows users to utilise a fully automated functional annotation pipeline directly from their internet browser. This would eliminate the need to install CHAP and its dependencies on a local workstation or computer cluster, a process that may be challenging for users unfamiliar with compiling software written in the C++ language. The

feasibility of online user interfaces for MD-based applications has been demonstrated by e.g. CHARMM-GUI,[341] a tool that generates complex input topologies for MD simulations. Realising the functional annotation protocol described in Chapter 3 would nonetheless be challenging, as it requires a server with scalable high performance computing capabilities. However, this could potentially be achieved on the basis of a cloud computing platform[342] equipped with modern general-purpose GPU hardware.

## 7.2 Hydrophobic Gates in Biological Ion Channels

Two pLGICs, the cation-selective 5-$HT_3$ receptor and the anion-selective glycine receptor, were used in Chapter 4 to demonstrate the utility of CHAP. Four structures of the 5-$HT_3$ receptor were functionally annotated on the basis of equilibrium MD simulations. This included the first X-ray structure of this channel (PDB ID: 4PIR), the functional status of which had originally been uncertain.[255] In addition, five structures of the glycine receptor were annotated using CHAP, yielding results in good agreement with the pharmacological properties of these structures.

Simulations of these two receptors were also used to assess the robustness of the simulation protocol introduced in Chapter 3. It was demonstrated that harmonic position restraints applied to the C$\alpha$ atoms of the protein ensure that the conformation of a simulated ion channel does not deviate substantially from that of an experimentally determined starting structure. Importantly, the restriction of backbone flexibility resulting from these restraints does not alter the liquid-vapour equilibrium of water inside the channel pore. Furthermore, significant savings in computational cost were achieved by simulating only a reduced subset of a channel protein, such as its pore-lining $\alpha$-helices. It was shown that the omission of protein domains distal to the putative location of a hydrophobic gate does not affect the behaviour of water confined inside it.

In addition to analysing two particular channels in detail, a high-throughput study was carried out to assess the prevalence of hydrophobic gates across all structurally

characterised ion channels. CHAP was used to show that pore de-wetting arises in ∼20 % of the ion channel structural proteome and may occur in hydrophobic pores with a radius of up to ∼0.4 nm. These results emphasise the importance of the hydrophobic gating mechanism and demonstrate the need for a tool with the functionality of CHAP that is able to reliably identify hydrophobic gates in novel channel structures.

## 7.3 The Influence of Induced Polarisation

An important open question in the literature on hydrophobic gating concerned the influence of induced polarisation. This was addressed in Chapter 5 through polarisable MD simulations of the 5-$HT_3$ receptor using the AMOEBA and the CHARMM Drude next generation force fields. It was shown that water molecules experience a remarkable reduction of their dipole moment as they pass through the hydrophobic gate of this channel, which is comparable in magnitude to the difference between liquid water and water vapour.

A comparison between these polarisable force fields and a large number of additive water models demonstrated qualitative agreement in predicting the hydration state of the 5-$HT_3$ receptor pore. However, for the open conformation of the channel, different water models predicted different degrees of de-wetting. The AMOEBA model favoured the liquid state, while the CHARMM Drude force field and most additive water models predicted an intermediate degree of wetting and frequent liquid-vapour transitions.

While the qualitative agreement between the AMOEBA and CHARMM Drude force fields in predicting a significant change of the water dipole moment in a hydrophobic gate is encouraging, future research is needed to precisely quantify the magnitude of this effect. Recent developments in linearly scaling density functional theory now permit electronic structure calculations on systems comprising >10 000 atoms[343] and may thus provide an avenue for directly determining molecular dipole moments of confined water without reliance on the transferability assumption inherent to MD force fields. An

alternative approach may be electronically coarse-grained MD based on the quantum Drude oscillator model.[344,345] Quantum Drude oscillators have been successfully applied to water,[346,347] but unfortunately parameters for biomolecules are not available yet.

Electronic polarisation also influences the behaviour of ions. As was demonstrated through potential of mean force calculations, polarisable ions experience a more ragged free energy landscape as they pass through the 5-$HT_3$ receptor pore compared to what is predicted by additive force fields. Polarisation was shown to increase the interaction of $Na^+$ ions with negatively charged amino acids, while $Cl^-$ ions were seen to substitute a part of their hydration shell for tighter interactions with uncharged protein residues. These results suggest that polarisable ions will experience increased dynamical friction when permeating an ion channel, which in turn would result in a lowered estimate of channel conductivity in comparison to additive MD simulations.

## 7.4 Electrowetting of Hydrophobic Gates

The role of a transmembrane potential in hydrophobic gating was explored in Chapter 6. It was shown that an external electric field can induce the wetting of the hydrophobic gate in the 5-$HT_3$ receptor. However, a supraphysiological voltage is required to achieve a significant degree of hydration. At physiological voltage levels, pore de-wetting effectively prohibits ion permeation and thus functionally closes the 5-$HT_3$ receptor. The electrowetting effect is therefore unlikely to be important for this channel.

Quantitatively, the wetting response can be described by a thermodynamic model based on surface tensions, the volumetric energy contribution of the electric field, and the effect of charged amino acids on the protein. It was shown through Poisson-Boltzmann electrostatics calculations that an unequal distribution of charged residues can cause a net electric field intrinsic to the protein. Depending on its orientation, this intrinsic field may increase or decrease the effect of the external field. This makes the wetting response asymmetric and in principle could be exploited to rectify mass transport through

hydrophobic nanopores.

Moreover, the electric field was also used to determine the kinetic constants characterising the phase transitions of water confined inside the 5-$HT_3$ receptor. It was established that when both states are equally probable, liquid-vapour oscillations occur on a time scale of ∼5 ns. This characteristic time scale was largely independent of the water model and was similar for the closed and open conformations of the 5-$HT_3$ receptor. When the electric field was adjusted to bias the system towards the liquid or vapour states, the life times of both states had inverse dependencies on their free energy difference. Consequently, it was shown that hydrophobic gates can be reliably identified through equilibrium MD simulations in the absence of an electric field and that a simulation time of ∼50 ns will typically be sufficient.

## 7.5 Conclusion

This thesis is concerned with the detection of hydrophobic gates in newly determined ion channel structures. Through atomistic MD simulation, it was demonstrated that hydrophobic gating is a relevant mechanism for a significant fraction of the ion channel structural proteome. The robustness of this finding was put to the test in two important ways. Firstly, advanced force fields were employed to explore the impact of induced electronic polarisation on the behaviour of water confined in ion channel pores. Secondly, it was shown that hydrophobic gates are not prone to electrowetting and reliably prevent the permeation of ions under the influence of a physiological transmembrane voltage. These insights were incorporated into the development of a new software suite capable of identifying hydrophobic gates, which has been made freely available under the name Channel Annotation Package — or CHAP for short. It is hoped that this new programme will prove itself a valuable tool for the structural biology community in its ongoing quest to elucidate the relationship between ion channel structure and function.

# A

# Polarisable Force Field Benchmark

| | Units | CHARMM36m | AMOEBA | CHARMM Drude |
|---|---|---|---|---|
| Implementation | — | GROMACS 2018 | OpenMM 7.1.1 | OpenMM 7.3.0 |
| System Size | $1 \times 10^3$ | 75.0 | 75.0 | 100.0 |
| Performance | $\mathrm{ns\,d^{-1}}$ | 80.0 | 7.4 | 29.9 |
| Performance | $\mathrm{ns\,h^{-1}}$ | 3.3 | 0.3 | 1.3 |

**Table A.1:** Typical performance of the polarisable AMOEBA and CHARMM Drude force fields compared to the additive CHARMM36m force field in simulations of the M2 helix bundle of the 5-$HT_3$ receptor. Polarisable simulations employed the AMOEBA14 and SWM4-NDP water models respectively, while additive simulations used the mTIP3P water model. The simulation system contains approximately 30 % more dynamically integrated particles in the CHARMM Drude force field due to the inclusion of Drude oscillators. Virtual sites do not count towards the system size. Simulations were performed on the Joint Academic Data Science Endeavour (JADE) cluster, employing one Nvidia V100 GPU per node. Algorithmic parameters were identical to those described in Section 5.1. Note that different MD implementations were used for each set of simulations and the benchmark data can therefore only serve as an approximate guide rather than as a rigorous comparison.

# Bibliography


[1] M. Luckey. *Membrane Structural Biology: With Biochemical and Biophysical Foundations*. Cambridge University Press, 2014.

[2] B. Hille. *Ionic Channels of Excitable Membranes*. Sinauer Associates Incorporated, 1992.

[3] C. M. Armstrong and B. Hille. "Voltage-Gated Ion Channels and Electrical Excitability". In: *Neuron* 20.3 (1998), pp. 371–380.

[4] H. Betz. "Ligand-Gated Ion Channels in the Brain: The Amino Acid Receptor Superfamily". In: *Neuron* 5.4 (1990), pp. 383–392.

[5] J. Wu, A. H. Lewis, and J. Grandl. "Touch, Tension, and Transduction — the Function and Regulation of Piezo Ion Channels". In: *Trends in Biochemical Sciences* 42.1 (2017), pp. 57–71.

[6] E. S. Haswell, R. Phillips, and D. C. Rees. "Mechanosensitive Channels: What Can They Do and How Do They Do It?" In: *Structure* 19.10 (2011), pp. 1356–1369.

[7] G. B. Monshausen and E. S. Haswell. "A Force of Nature: Molecular Mechanisms of Mechanoperception in Plants". In: *Journal of Experimental Botany* 64.15 (2013), pp. 4663–4680.

[8] F. M. Ashcroft. *Ion Channels and Disease*. Academic Press, 1999.

[9] B. Roux. "Ion Channels and Ion Selectivity". In: *Essays in Biochemistry* 61.2 (2017), pp. 201–209.

[10] F. M. Ashcroft. "From Molecule to Malady". In: *Nature* 440–447.7083 (2006), p. 440.

[11] M. E. Curran, I. Splawski, K. W. Timothy, G. M. Vincen, E. D. Green, and M. T. Keating. "A Molecular Basis for Cardiac Arrhythmia: HERG Mutations Cause Long QT Syndrome". In: *Cell* 80.5 (1995), pp. 795–803.

[12] M. C. Sanguinetti and M. Tristani-Firouzi. "hERG Potassium Channels and Cardiac Arrhythmia". In: *Nature* 440.7083 (2006), pp. 463–469.

[13] D. L. Browne, S. T. Gancher, J. G. Nutt, E. R. P. Brunt, E. A. Smith, P. Kramer, and M. Litt. "Episodic Ataxia/Myokymia Syndrome is Associated with Point Mutations in the Human Potassium Channel Gene, KCNA1". In: *Nature Genetics* 8.2 (1994), pp. 136–140.

[14] M. H. Meisler and J. A. Kearney. "Sodium Channel Mutations in Epilepsy and Other Neurological Disorders". In: *The Journal of Clinical Investigation* 115.8 (2005), pp. 2010–2017.

[15] F. M. Ashcroft. “ATP-Sensitive Potassium Channelopathies: Focus on Insulin Secretion”. In: *The Journal of Clinical Investigation* 115.8 (2005), pp. 2047–2058.

[16] D. C. Gadsby, P. Vergani, and L. Csanády. “The ABC Protein Turned Chloride Channel Whose Failure Causes Cystic Fibrosis”. In: *Nature* 440.7083 (2006), pp. 477–483.

[17] S. K. Bagal, A. D. Brown, P. J. Cox, K. Omoto, R. M. Owen, D. C. Pryde, B. Sidders, S. E. Skerratt, E. B. Stevens, R. I. Storer, et al. “Ion Channels as Therapeutic Targets: A Drug Discovery Perspective”. In: *Journal of Medicinal Chemistry* 56.3 (2012), pp. 593–624.

[18] J. P. Overington, B. Al-Lazikani, and A. L. Hopkins. “How Many Drug Targets Are There?” In: *Nature Reviews Drug Discovery* 5.12 (2006), pp. 993–996.

[19] A. Mathie. “Ion Channels as Novel Therapeutic Targets in the Treatment of Pain”. In: *Journal of Pharmacy and Pharmacology* 62.9 (2010), pp. 1089–1095.

[20] N. P. Franks and W. R. Lieb. “Molecular and Cellular Mechanisms of General Anaesthesia”. In: *Nature* 367.6464 (1994), pp. 607–614.

[21] R. MacKinnon, S. L. Cohen, A. Kuo, A. Lee, and B. T. Chait. “Structural Conservation in Prokaryotic and Eukaryotic Potassium Channels”. In: *Science* 280.5360 (1998), pp. 106–109.

[22] D. A. Doyle, J. M. Cabral, R. A. Pfuetzner, A. Kuo, J. M. Gulbis, S. L. Cohen, B. T. Chait, and R. MacKinnon. “The Structure of the Potassium Channel: Molecular Basis of $K^+$ Conduction and Selectivity”. In: *Science* 280.5360 (1998), pp. 69–77.

[23] K. R. Vinothkumar and R. Henderson. “Structures of Membrane Proteins”. In: *Quarterly Reviews of Biophysics* 43.1 (2010), pp. 65–158.

[24] J.-J. Lacapere, E. Pebay-Peyroula, J.-M. Neumann, and C. Etchebest. “Determining Membrane Protein Structures: Still a Challenge!” In: *Trends in Biochemical Sciences* 32.6 (2007), pp. 259–270.

[25] E. P. Carpenter, K. Beis, A. D. Cameron, and S. Iwata. “Overcoming the Challenges of Membrane Protein Cystallography”. In: *Current Opinion in Structural Biology* 18.5 (2008), pp. 581–586.

[26] G. Klesse, S. Rao, M. S. P. Sansom, and S. J. Tucker. “CHAP: A Versatile Tool for the Structural and Functional Annotation of Ion Channel Pores”. In: *Journal of Molecular Biology* 431.17 (2019), pp. 3353–3365.

[27] X.-C. Bai, G. McMullan, and S. H. W. Scheres. “How Cryo-EM is Revolutionizing Structural Biology”. In: *Trends in Biochemical Sciences* 40.1 (2015), pp. 49–57.

[28] W. Kühlbrandt. “The Resolution Revolution”. In: *Science* 343.6178 (2014), pp. 1443–1444.

[29] L. A. Earl, V. Falconieri, J. L. S. Milne, and S. Subramaniam. “Cryo-EM: Beyond the Microscope”. In: *Current Opinion in Structural Biology* 46 (2017), pp. 71–78.

[30] Z. Yan, X.-c. Bai, C. Yan, J. Wu, Z. Li, T. Xie, W. Peng, C.-c. Yin, X. Li, S. H. Scheres, et al. "Structure of the Rabbit Ryanodine Receptor RyR1 at Near-Atomic Resolution". In: *Nature* 517.7532 (2015), pp. 50–55.

[31] K. Saotome, S. E. Murthy, J. M. Kefauver, T. Whitwam, A. Patapoutian, and A. B. Ward. "Structure of the Mechanically Activated Ion Channel Piezo1". In: *Nature* 554.7693 (2018), pp. 481–486.

[32] D. P. Tieleman, P. C. Biggin, G. R. Smith, and M. S. P. Sansom. "Simulation Approaches to Ion Channel Structure-Function Relationships". In: *Quarterly Reviews of Biophysics* 34.4 (2001), pp. 473–561.

[33] O. Beckstein, P. C. Biggin, P. Bond, J. N. Bright, C. Domene, A. Grottesi, J. Holyoake, and M. S. P. Sansom. "Ion Channel Gating: Insights via Molecular Simulations". In: *FEBS Letters* 555.1 (2003), pp. 85–90.

[34] W. L. Ash, M. R. Zlomislic, E. O. Oloo, and D. P. Tieleman. "Computer Simulations of Membrane Proteins". In: *Biochimica et Biophysica Acta — Biomembranes* 1666.1-2 (2004), pp. 158–189.

[35] B. Roux, T. Allen, S. Berneche, and W. Im. "Theoretical and Computational Models of Biological Ion Channels". In: *Quarterly Reviews of Biophysics* 37.1 (2004), pp. 15–103.

[36] J. L. Trick, P. Aryal, S. J. Tucker, and M. S. P. Sansom. "Molecular Simulation Studies of Hydrophobic Gating in Nanopores and Ion Channels". In: *Biochemical Society Transactions* 43.2 (2015), pp. 146–150.

[37] P. Aryal, M. S. P. Sansom, and S. J. Tucker. "Hydrophobic Gating in Ion Channels". In: *Journal of Molecular Biology* 427.1 (2015), pp. 121–130.

[38] S. Rao, C. I. Lynch, G. Klesse, G. E. Oakley, P. J. Stansfeld, S. J. Tucker, and M. S. P. Sansom. "Water and Hydrophobic Gates in Ion Channels and Nanopores". In: *Faraday Discussions* 209 (2018), pp. 231–247.

[39] R. Allen, S. Melchionna, and J.-P. Hansen. "Intermittent Permeation of Cylindrical Nanopores by Water". In: *Physical Review Letters* 89.17 (2002).

[40] R. Allen, J.-P. Hansen, and S. Melchionna. "Molecular Dynamics Investigation of Water Permeation Through Nanopores". In: *The Journal of Chemical Physics* 119.7 (2003), pp. 3905–3919.

[41] O. Beckstein and M. S. P. Sansom. "Liquid-Vapor Oscillations of Water in Hydrophobic Nanopores". In: *Proceedings of the National Academy of Sciences* 100.12 (2003), pp. 7063–7068.

[42] G. Hummer, J. C. Rasaiah, and J. P. Noworyta. "Water Conduction Through the Hydrophobic Channel of a Carbon Nanotube". In: *Nature* 414.6860 (2001), pp. 188–190.

[43] A. Waghe, J. C. Rasaiah, and G. Hummer. "Filling and Emptying Kinetics of Carbon Nanotubes in Water". In: *The Journal of Chemical Physics* 117.23 (2002), pp. 10789–10795.

[44] O. Beckstein and M. S. P. Sansom. "The Influence of Geometry, Surface Character, and Flexibility on the Permeation of Ions and Water Through Biological Pores". In: *Physical Biology* 1.1 (2004), pp. 42–52.

[45] O. Beckstein, P. C. Biggin, and M. S. P. Sansom. "A Hydrophobic Gating Mechanism for Nanopores". In: *The Journal of Physical Chemistry B* 105.51 (2001), pp. 12902–12905.

[46] R. Schmid, A. M. Miah, and V. N. Sapunov. "A New Table of the Thermodynamic Quantities of Ionic Hydration: Values and Some Applications (Enthalpy-Entropy Compensation and Born Radii)". In: *Physical Chemistry Chemical Physics* 2.1 (2000), pp. 97–102.

[47] J. Dzubiella, R. J. Allen, and J.-P. Hansen. "Electric Field-Controlled Water Permeation Coupled to Ion Transport Through a Nanopore". In: *The Journal of Chemical Physics* 120.11 (2004), pp. 5001–5004.

[48] J. Dzubiella and J.-P. Hansen. "Electric-Field-Controlled Water and Ion Permeation of a Hydrophobic Nanopore". In: *The Journal of Chemical Physics* 122.23 (2005), p. 234706.

[49] J. Li, X. Gong, H. Lu, D. Li, H. Fang, and R. Zhou. "Electrostatic Gating of a Nanometer Water Channel". In: *Proceedings of the National Academy of Sciences* 104.10 (2007), pp. 3687–3692.

[50] P. Aryal, F. Abd-Wahab, G. Bucci, M. S. P. Sansom, and S. J. Tucker. "A Hydrophobic Barrier Deep Within the Inner Pore of the TWIK-1 K2P Potassium Channel". In: *Nature Communications* 5 (2014), p. 4377.

[51] J. P. Birkner, B. Poolman, and A. Koçer. "Hydrophobic Gating of Mechanosensitive Channel of Large Conductance Evidenced by Single-Subunit Resolution". In: *Proceedings of the National Academy of Sciences* 109.32 (2012), pp. 12944–12949.

[52] R. Roth, D. Gillespie, W. Nonner, and R. E. Eisenberg. "Bubbles, Gating, and Anesthetics in Ion Channels". In: *Biophysical Journal* 94.11 (2008), pp. 4282–4298.

[53] M. O. Jensen, D. W. Borhani, K. Lindorff-Larsen, P. Maragakis, V. Jogini, M. P. Eastwood, R. O. Dror, and D. E. Shaw. "Principles of Conduction and Hydrophobic Gating in $K^+$ Channels". In: *Proceedings of the National Academy of Sciences* 107.13 (2010), pp. 5833–5838.

[54] L. Sauguet, A. Shahsavar, and M. Delarue. "Crystallographic Studies of Pharmacological Sites in Pentameric Ligand-Gated Ion Channels". In: *Biochimica et Biophysica Acta — General Subjects* 1850.3 (2015), pp. 511–523.

[55] R. B. Bass, P. Strop, M. Barclay, and D. C. Rees. "Crystal Structure of Escherichia Coli MscS, a Voltage-Modulated and Mechanosensitive Channel". In: *Science* 298.5598 (2002), pp. 1582–1587.

[56] A. Anishkin and S. Sukharev. "Water Dynamics and Dewetting Transitions in the Small Mechanosensitive Channel MscS". In: *Biophysical Journal* 86.5 (2004), pp. 2883–2895.

[57] M. Sotomayor and K. Schulten. "Molecular Dynamics Study of Gating in the Mechanosensitive Channel of Small Conductance MscS". In: *Biophysical Journal* 87.5 (2004), pp. 3050–3065.

[58] S. Miller, W. Bartlett, S. Chandrasekaran, S. Simpson, M. Edwards, and I. R. Booth. "Domain Organization of the MscS Mechanosensitive Channel of Escherichia Coli". In: *The EMBO Journal* 22.1 (2003), pp. 36–46.

[59] W. Wang, S. S. Black, M. D. Edwards, S. Miller, E. L. Morrison, W. Bartlett, C. Dong, J. H. Naismith, and I. R. Booth. "The Structure of an Open Form of an E. Coli Mechanosensitive Channel at 3.45 Å Resolution". In: *Science* 321.5893 (2008), pp. 1179–1183.

[60] A. Anishkin, B. Akitake, K. Kamaraju, C. S. Chiang, and S. Sukharev. "Hydration Properties of Mechanosensitive Channel Pores Define the Energetics of Gating". In: *Journal of Physics: Condensed Matter* 22.45 (2010), p. 454120.

[61] S. A. Spronk, D. E. Elmore, and D. A. Dougherty. "Voltage-Dependent Hydration and Conduction Properties of the Hydrophobic Pore of the Mechanosensitive Channel of Small Conductance". In: *Biophysical Journal* 90.10 (2006), pp. 3555–3569.

[62] A. Miyazawa, Y. Fujiyoshi, and N. Unwin. "Structure and Gating Mechanism of the Acetylcholine Receptor Pore". In: *Nature* 423.6943 (2003), pp. 949–955.

[63] O. Beckstein and M. S. P. Sansom. "A Hydrophobic Gate in an Ion Channel: The Closed State of the Nicotinic Acetylcholine Receptor". In: *Physical Biology* 3.2 (2006), pp. 147–159.

[64] F. Revah, D. Bertrand, J.-L. Galzi, A. Devillers-Thiery, C. Mulle, N. Hussy, S. Bertrand, M. Ballivet, and J.-P. Changeux. "Mutations in the Channel Domain Alter Desensitization of a Neuronal Nicotinic Receptor". In: *Nature* 353.6347 (1991), pp. 846–849.

[65] G. N. Filatov and M. M. White. "The Role of Conserved Leucines in the M2 Domain of the Acetylcholine Receptor in Channel Gating". In: *Molecular Pharmacology* 48.3 (1995), pp. 379–384.

[66] C. Labarca, M. W. Nowak, H. Zhang, L. Tang, P. Deshpande, and H. A. Lester. "Channel Gating Governed Symmetrically by Conserved Leucine Residues in the M2 Domain of Nicotinic Receptors". In: *Nature* 376.6540 (1995), pp. 514–516.

[67] P. V. Plazas, M. J. De Rosa, M. E. Gomez-Casati, M. Verbitsky, N. Weisstaub, E. Katz, C. Bouzat, and A. B. Elgoyhen. "Key Roles of Hydrophobic Rings of TM2 in Gating of the $\alpha 9\alpha 10$ Nicotinic Cholinergic Receptor". In: *British Journal of Pharmacology* 145.7 (2005), pp. 963–974.

[68] R. J. C. Hilf and R. Dutzler. "X-Ray Structure of a Prokaryotic Pentameric Ligand-Gated Ion Channel". In: *Nature* 452.7185 (2008), pp. 375–379.

[69] R. J. C. Hilf and R. Dutzler. "Structure of a Potentially Open State of a Proton-Activated Pentameric Ligand-Gated Ion Channel". In: *Nature* 457.7225 (2009), pp. 115–118.

[70] N. Bocquet, H. Nury, M. Baaden, C. Le Poupon, J.-P. Changeux, M. Delarue, and P.-J. Corringer. "X-Ray Structure of a Pentameric Ligand-Gated Ion Channel in an Apparently Open Conformation". In: *Nature* 457.7225 (2009), pp. 111–114.

[71] F. Zhu and G. Hummer. "Pore Opening and Closing of a Pentameric Ligand-Gated Ion Channel". In: *Proceedings of the National Academy of Sciences* 107.46 (2010), pp. 19814–19819.

[72] F. Zhu and G. Hummer. "Drying Transition in the Hydrophobic Gate of the GLIC Channel Blocks Ion Conduction". In: *Biophysical Journal* 103.2 (2012), pp. 219–227.

[73] Z. Jia, M. Yazdani, G. Zhang, J. Cui, and J. Chen. "Hydrophobic Gating in BK Channels". In: *Nature Communications* 9.1 (2018), p. 3408.

[74] M. A. Kasimova, A. Yazici, Y. Yudin, D. Granata, M. L. Klein, T. Rohacs, and V. Carnevale. "Ion Channel Sensing: Are Fluctuations the Crux of the Matter?" In: *The Journal of Physical Chemistry Letters* 9.6 (2018), pp. 1260–1264.

[75] C. Neale, N. Chakrabarti, P. Pomorski, E. F. Pai, and R. Pomès. "Hydrophobic Gating of Ion Permeation in Magnesium Channel CorA". In: *PLOS Computational Biology* 11.7 (2015), e1004303.

[76] M. Yamashita, P. S.-W. Yeung, C. E. Ing, B. A. McNally, R. Pomès, and M. Prakriya. "STIM1 Activates CRAC Channels Through Rotation of the Pore Helix to Open a Hydrophobic Gate". In: *Nature Communications* 8 (2017), p. 14512.

[77] S. Rao, G. Klesse, P. J. Stansfeld, S. J. Tucker, and M. S. P. Sansom. "A BEST Example of Channel Structure Annotation by Molecular Simulation". In: *Channels* 11.4 (2017), pp. 347–353.

[78] J. L. Trick, S. Chelvaniththilan, G. Klesse, P. Aryal, E. J. Wallace, S. J. Tucker, and M. S. P. Sansom. "Functional Annotation of Ion Channel Structures by Molecular Simulation". In: *Structure* 24.12 (2016), pp. 2207–2216.

[79] J. L. Trick, C. Song, E. J. Wallace, and M. S. P. Sansom. "Voltage Gating of a Biomimetic Nanopore: Electrowetting of a Hydrophobic Barrier". In: *ACS Nano* 11.2 (2017), pp. 1840–1847.

[80] F. C. Chatelain, D. Bichet, D. Douguet, S. Feliciangeli, S. Bendahhou, M. Reichold, R. Warth, J. Barhanin, and F. Lesage. "TWIK1, a Unique Background Channel with Variable Ion Selectivity". In: *Proceedings of the National Academy of Sciences* 109.14 (2012), pp. 5499–5504.

[81] W. Zheng, R. Hu, R. Cai, L. Hofmann, Q. Hu, M. Fatehi, W. Long, T. Kong, J. Tang, P. Light, V. Flockerzi, Y. Cao, and X.-Z. Chen. "Identification and Characterization of Hydrophobic Gate Residues in TRP Channels". In: *The FASEB Journal* 32.2 (2018), pp. 639–653.

[82] W. Zheng, E. O. Gracheva, and S. N. Bagriantsev. "A Hydrophobic Gate in the Inner Pore Helix is the Major Determinant of Inactivation in Mechanosensitive Piezo Channels". In: *eLife* 8 (2019), e44003.

[83] M. R. Powell, L. Cleary, M. Davenport, K. J. Shea, and Z. S. Siwy. "Electric-Field-Induced Wetting and Dewetting in Single Hydrophobic Nanopores". In: *Nature Nanotechnology* 6.12 (2011), pp. 798–802.

[84] T. von Grotthuß. *Mémoire sur la décomposition de l'eau et des corps qu'elle tient en dissolution à l'aide de l'électricité galvanique*. 1805.

[85] S. Cukierman. "Et Tu, Grotthuss! and Other Unfinished Stories". In: *Biochimica et Biophysica Acta — Bioenergetics* 1757.8 (2006), pp. 876–885.

[86] S. Vaitheeswaran, J. C. Rasaiah, and G. Hummer. "Electric Field and Temperature Effects on Water in the Narrow Nonpolar Pores of Carbon Nanotubes". In: *The Journal of Chemical Physics* 121.16 (2004), pp. 7955–7965.

[87] S. Gupta, S. Chakraborty, R. Vij, and A. Auerbach. "A Mechanism for Acetylcholine Receptor Gating Based on Structure, Coupling, Phi, and Flip". In: *The Journal of General Physiology* 149.1 (2017), pp. 85–103.

[88] A. Jha, P. Purohit, and A. Auerbach. "Energy and Structure of the M2 Helix in Acetylcholine Receptor-Channel Gating". In: *Biophysical Journal* 96.10 (2009), pp. 4075–4084.

[89] A. Auerbach. "The Gating Isomerization of Neuromuscular Acetylcholine Receptors". In: *The Journal of Physiology* 588.4 (2010), pp. 573–586.

[90] B. Roux. "Theoretical and Computational Models of Ion Channels". In: *Current Opinion in Structural Biology* 12.2 (2002), pp. 182–189.

[91] B. Roux. "The Calculation of the Potential of Mean Force Using Computer Simulations". In: *Computer Physics Communications* 91.1-3 (1995), pp. 275–282.

[92] B. Roux. "The Membrane Potential and Its Representation by a Constant Electric Field in Computer Simulations". In: *Biophysical Journal* 95.9 (2008), pp. 4205–4216.

[93] C. Kutzner, H. Grubmüller, B. L. De Groot, and U. Zachariae. "Computational Electrophysiology: the Molecular Dynamics of Ion Channel Permeation and Selectivity in Atomistic Detail". In: *Biophysical Journal* 101.4 (2011), pp. 809–817.

[94] J. E. Stone, D. J. Hardy, I. S. Ufimtsev, and K. Schulten. "GPU-Accelerated Molecular Modeling Coming of Age". In: *Journal of Molecular Graphics and Modelling* 29.2 (2010), pp. 116–125.

[95] C. Kutzner, S. Páll, M. Fechner, A. Esztermann, B. L. de Groot, and H. Grubmüller. "Best Bang for Your Buck: GPU Nodes for GROMACS Biomolecular Simulations". In: *Journal of Computational Chemistry* 36.26 (2015), pp. 1990–2008.

[96] T.-S. Lee, D. S. Cerutti, D. Mermelstein, C. Lin, S. LeGrand, T. J. Giese, A. Roitberg, D. A. Case, R. C. Walker, and D. M. York. "GPU-Accelerated Molecular Dynamics and Free Energy Methods in Amber18: Performance Enhancements and New Features". In: *Journal of Chemical Information and Modeling* 58.10 (2018), pp. 2043–2050.

[97] C. Kutzner, S. Páll, M. Fechner, A. Esztermann, B. L. de Groot, and H. Grubmüller. *More Bang for Your Buck: Improved Use of GPU Nodes for GROMACS 2018*. 2019. arXiv: `1903.05918`.

[98] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, and E. Lindahl. "GROMACS: High Performance Molecular Simulations Through Multi-level Parallelism from Laptops to Supercomputers". In: *SoftwareX* 1 (2015), pp. 19–25.

[99] J. W. Ponder, C. Wu, P. Ren, V. S. Pande, J. D. Chodera, M. J. Schnieders, I. Haque, D. L. Mobley, D. S. Lambrecht, R. A. DiStasio Jr, et al. "Current Status of the AMOEBA Polarizable Force Field". In: *The Journal of Physical Chemistry B* 114.8 (2010), pp. 2549–2564.

[100] J. A. Lemkul, B. Roux, D. van der Spoel, and A. D. MacKerell Jr. "Implementation of Extended Lagrangian Dynamics in GROMACS for Polarizable Simulations Using the Classical Drude Oscillator Model". In: *Journal of Computational Chemistry* 36.19 (2015), pp. 1473–1479.

[101] X. Peng, Y. Zhang, H. Chu, Y. Li, D. Zhang, L. Cao, and G. Li. "Accurate Evaluation of Ion Conductivity of the Gramicidin A Channel Using a Polarizable Force Field Without Any Corrections". In: *Journal of Chemical Theory and Computation* 12.6 (2016), pp. 2973–2982.

[102] C. J. Illingworth, S. Furini, and C. Domene. "Computational Studies on Polarization Effects and Selectivity in $K^+$ Channels". In: *Journal of Chemical Theory and Computation* 6.12 (2010), pp. 3780–3792.

[103] S. Rao, G. Klesse, P. J. Stansfeld, S. J. Tucker, and M. S. P. Sansom. "A Heuristic Derived From Analysis of the Ion Channel Structural Proteome Permits the Rapid Identification of Hydrophobic Gates". In: *Proceedings of the National Academy of Sciences* (2019), pp. 13989–13995.

[104] J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser, and C. Simmerling. "ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB". In: *Journal of Chemical Theory and Computation* 11.8 (2015), pp. 3696–3713.

[105] R. B. Best, X. Zhu, J. Shim, P. E. M. Lopes, J. Mittal, M. Feig, and A. D. MacKerell Jr. "Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone $\phi$, $\psi$ and Side-Chain $\chi_1$ and $\chi_2$ Dihedral Angles". In: *Journal of Chemical Theory and Computation* 8.9 (2012), pp. 3257–3273.

[106] M. M. Reif, P. H. Hünenberger, and C. Oostenbrink. "New Interaction Parameters for Charged Amino Acid Side Chains in the GROMOS Force Field". In: *Journal of Chemical Theory and Computation* 8.10 (2012), pp. 3705–3723.

[107] W. L. Jorgensen and C. J. Swenson. "Optimized Intermolecular Potential Functions for Amides and Peptides. Structure and Properties of Liquid Amides". In: *Journal of the American Chemical Society* 107.3 (1985), pp. 569–578.

[108] X. Periole and S.-J. Marrink. "The Martini Coarse-Grained Force Field". In: *Biomolecular Simulations*. Springer, 2013, pp. 533–565.

[109] J. Huang, S. Rauscher, G. Nawrocki, T. Ran, M. Feig, B. L. de Groot, H. Grubmüller, and A. D. MacKerell Jr. "CHARMM36m: An Improved Force Field for Folded and Intrinsically Disordered Proteins". In: *Nature Methods* 14.1 (2017), pp. 71–73.

[110] A. Sandoval-Perez, K. Pluhackova, and R. A. Böckmann. "Critical Comparison of Biomembrane Force Fields: Protein-Lipid Interactions at the Membrane Interface". In: *Journal of Chemical Theory and Computation* 13.5 (2017), pp. 2310–2321.

[111] W. C. Still, A. Tempczyk, R. C. Hawley, and T. Hendrickson. "Semianalytical Treatment of Solvation for Molecular Mechanics and Dynamics". In: *Journal of the American Chemical Society* 112.16 (1990), pp. 6127–6129.

[112] B. Z. Lu, Y. C. Zhou, M. J. Holst, and J. A. McCammon. "Recent Progress in Numerical Methods for the Poisson-Boltzmann Equation in Biophysical Applications". In: *Communications in Computer Physics* 3.5 (2008), pp. 973–1009.

[113] M. W. Mahoney and W. L. Jorgensen. "A Five-Site Model for Liquid Water and the Reproduction of the Density Anomaly by Rigid, Nonpolarizable Potential Functions". In: *The Journal of Chemical Physics* 112.20 (2000), pp. 8910–8922.

[114] H. Nada and J. P. J. M. van der Eerden. "An Intermolecular Potential Model for the Simulation of Ice and Water Near the Melting Point: A Six-Site Model of $H_2O$". In: *The Journal of Chemical Physics* 118.16 (2003), pp. 7401–7413.

[115] B. Guillot. "A Reappraisal of What We Have Learnt During Three Decades of Computer Simulations on Water". In: *Journal of Molecular Liquids* 101.1-3 (2002), pp. 219–260.

[116] SklogWiki. *SklogWiki Water Category Overview*. 2019. URL: `http://www.sklogwiki.org/SklogWiki/index.php/Category:Water`.

[117] A. V. Onufriev and S. Izadi. "Water Models for Biomolecular Simulations". In: *Wiley Interdisciplinary Reviews: Computational Molecular Science* 8.2 (2018), e1347.

[118] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. "Comparison of Simple Potential Functions for Simulating Liquid Water". In: *The Journal of Chemical Physics* 79.2 (1983), pp. 926–935.

[119] W. L. Jorgensen and J. D. Madura. "Temperature and Size Dependence for Monte Carlo Simulations of TIP4P Water". In: *Molecular Physics* 56.6 (1985), pp. 1381–1392.

[120] S. R. Durell, B. R. Brooks, and A. Ben-Naim. "Solvent-Induced Forces Between Two Hydrophilic Groups". In: *The Journal of Physical Chemistry* 98.8 (1994), pp. 2198–2202.

[121] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, and J. Hermans. "Interaction Models for Water in Relation to Protein Hydration". In: *Intermolecular Forces*. Springer, 1981, pp. 331–342.

[122] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma. "The Missing Term in Effective Pair Potentials". In: *Journal of Physical Chemistry* 91.24 (1987), pp. 6269–6271.

[123] J. L. F. Abascal and C. Vega. "A General Purpose Model for the Condensed Phases of Water: TIP4P/2005". In: *The Journal of Chemical Physics* 123.23 (2005), p. 234505.

[124] S. Izadi, R. Anandakrishnan, and A. V. Onufriev. "Building Water Models: A Different Approach". In: *The Journal of Physical Chemistry Letters* 5.21 (2014), pp. 3863–3871.

[125] D. J. Price and C. L. Brooks III. "A modified TIP3P Water Potential for Simulation With Ewald Summation". In: *The Journal of Chemical Physics* 121.20 (2004), pp. 10096–10103.

[126] H. W. Horn, W. C. Swope, J. W. Pitera, J. D. Madura, T. J. Dick, G. L. Hura, and T. Head-Gordon. "Development of an Improved Four-Site Water Model for Biomolecular Simulations: TIP4P-Ew". In: *The Journal of Chemical Physics* 120.20 (2004), pp. 9665–9678.

[127] U. W. Schmitt and G. A. Voth. "The Computer Simulation of Proton Transport in Water". In: *The Journal of Chemical Physics* 111.20 (1999), pp. 9361–9381.

[128] C. P. Lawrence and J. L. Skinner. "Flexible TIP4P Model for Molecular Dynamics Simulation of Liquid Water". In: *Chemical Physics Letters* 372.5-6 (2003), pp. 842–847.

[129] C. Vega and E. De Miguel. "Surface Tension of the Most Popular Models of Water by Using the Test-Area Simulation Method". In: *The Journal of Chemical Physics* 126.15 (2007), p. 154707.

[130] C. Vega and J. L. F. Abascal. "Simulating Water With Rigid Non-Polarizable Models: A General Perspective". In: *Physical Chemistry Chemical Physics* 13.44 (2011), pp. 19663–19688.

[131] H. Zhang, C. Yin, Y. Jiang, and D. van der Spoel. "Force Field Benchmark of Amino Acids: I. Hydration and Diffusion in Different Water Models". In: *Journal of Chemical Information and Modeling* 58.5 (2018), pp. 1037–1052.

[132] S. Izadi and A. V. Onufriev. "Accuracy Limit of Rigid 3-Point Water Models". In: *The Journal of Chemical Physics* 145.7 (2016), p. 74501.

[133] R. W. Hockney, S. P. Goel, and J. W. Eastwood. "Quiet High-Resolution Computer Models of a Plasma". In: *Journal of Computational Physics* 14.2 (1974), pp. 148–158.

[134] H. J. C. Berendsen and W. F. Van Gunsteren. "Practical Algorithms for Dynamic Simulations". In: *Molecular-Dynamics Simulation of Statistical-Mechanical Systems* (1986), pp. 43–65.

[135] W. F. Van Gunsteren and M. Karplus. "Effect of Constraints on the Dynamics of Macromolecules". In: *Macromolecules* 15.6 (1982), pp. 1528–1544.

[136] J.-P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen. "Numerical Integration of the Cartesian Equations of Motion of a System With Constraints: Molecular Dynamics of $n$-Alkanes". In: *Journal of Computational Physics* 23.3 (1977), pp. 327–341.

[137] S. Miyamoto and P. A. Kollman. "SETTLE: An Analytical Version of the SHAKE and RATTLE Algorithm for Rigid Water Models". In: *Journal of Computational Chemistry* 13.8 (1992), pp. 952–962.

[138] B. Hess, H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije. "LINCS: A Linear Constraint Solver for Molecular Simulations". In: *Journal of Computational Chemistry* 18.12 (1997), pp. 1463–1472.

[139] B. Hess. "P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation". In: *Journal of Chemical Theory and Computation* 4.1 (2008), pp. 116–122.

[140] H. C. Andersen. "Molecular Dynamics Simulations at Constant Pressure and/or Temperature". In: *The Journal of Chemical Physics* 72.4 (1980), pp. 2384–2393.

[141] J. E. Basconi and M. R. Shirts. "Effects of Temperature Control Algorithms on Transport Properties and Kinetics in Molecular Dynamics Simulations". In: *Journal of Chemical Theory and Computation* 9.7 (2013), pp. 2887–2899.

[142] G. Bussi, D. Donadio, and M. Parrinello. "Canonical Sampling Through Velocity Rescaling". In: *The Journal of Chemical Physics* 126.1 (2007), p. 14101.

[143] M. Parrinello and A. Rahman. "Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method". In: *Journal of Applied Physics* 52.12 (1981), pp. 7182–7190.

[144] S. Nosé and M. L. Klein. "Constant Pressure Molecular Dynamics for Molecular Systems". In: *Molecular Physics* 50.5 (1983), pp. 1055–1076.

[145] K.-H. Chow and D. M. Ferguson. "Isothermal-Isobaric Molecular Dynamics Simulations with Monte Carlo Volume Sampling". In: *Computer Physics Communications* 91.1-3 (1995), pp. 283–289.

[146] J. Åqvist, P. Wennerström, M. Nervall, S. Bjelic, and B. O. Brandsdal. "Molecular Dynamics Simulations of Water and Biomolecules with a Monte Carlo Constant Pressure Algorithm". In: *Chemical Physics Letters* 384.4-6 (2004), pp. 288–294.

[147] M. P. Allen and D. J. Tildesley. *Computer Simulation of Liquids*. Oxford University Press, 2017.

[148] P. P. Ewald. "Die Berechnung optischer und elektrostatischer Gitterpotentiale". In: *Annalen der Physik* 369.3 (1921), pp. 253–287.

[149] M. Deserno and C. Holm. "How to Mesh up Ewald Sums. I. A Theoretical and Numerical Comparison of Various Particle Mesh Routines". In: *The Journal of Chemical Physics* 109.18 (1998), pp. 7678–7693.

[150] J. W. Perram, H. G. Petersen, and S. W. De Leeuw. "An Algorithm for the Simulation of Condensed Matter Which Grows as the 3/2 Power of the Number of Particles". In: *Molecular Physics* 65.4 (1988), pp. 875–893.

[151] T. Darden, D. York, and L. Pedersen. "Particle Mesh Ewald: An $N \log N$ Method for Ewald Sums in Large Systems". In: *The Journal of Chemical Physics* 98.12 (1993), pp. 10089–10092.

[152] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen. "A Smooth Particle Mesh Ewald Method". In: *The Journal of Chemical Physics* 103.19 (1995), pp. 8577–8593.

[153] M. R. Shirts, D. L. Mobley, J. D. Chodera, and V. S. Pande. "Accurate and Efficient Corrections for Missing Dispersion Interactions in Molecular Simulations". In: *The Journal of Physical Chemistry B* 111.45 (2007), pp. 13052–13063.

[154] P. S. Nerenberg and T. Head-Gordon. "New Developments in Force Fields for Biomolecular Simulations". In: *Current Opinion in Structural Biology* 49 (2018), pp. 129–138.

[155] C. M. Baker. "Polarizable Force Fields for Molecular Dynamics Simulations of Biomolecules". In: *Wiley Interdisciplinary Reviews: Computational Molecular Science* 5.2 (2015), pp. 241–254.

[156] P. J. Van Maaren and D. Van Der Spoel. "Molecular Dynamics Simulations of Water With Novel Shell-Model Potentials". In: *The Journal of Physical Chemistry B* 105.13 (2001), pp. 2618–2626.

[157] A.-P. E. Kunz and W. F. van Gunsteren. "Development of a Nonlinear Classical Polarization Model for Fiquid Water and Aqueous Solutions: COS/D". In: *The Journal of Physical Chemistry A* 113.43 (2009), pp. 11570–11579.

[158] S. J. Stuart and B. J. Berne. "Effects of Polarizability on the Hydration of the Chloride Ion". In: *The Journal of Physical Chemistry* 100.29 (1996), pp. 11934–11943.

[159] C. Schröder and O. Steinhauser. "Simulating Polarizable Molecular Ionic Liquids With Drude Oscillators". In: *The Journal of Chemical Physics* 133.15 (2010), p. 154511.

[160] J. A. Lemkul, J. Huang, B. Roux, and A. D. MacKerell Jr. "An Empirical Polarizable Force Field Based on the Classical Drude Oscillator Model: Development History and Recent Applications". In: *Chemical Reviews* 116.9 (2016), pp. 4983–5013.

[161] G. Lamoureux and B. Roux. "Modeling Induced Polarization With Classical Drude Oscillators: Theory and Molecular Dynamics Simulation Algorithm". In: *The Journal of Chemical Physics* 119.6 (2003), pp. 3025–3039.

[162] W. Jiang, D. J. Hardy, J. C. Phillips, A. D. MacKerell Jr, K. Schulten, and B. Roux. "High-Performance Scalable Molecular Dynamics Simulations of a Polarizable Force Field Based on Classical Drude Oscillators in NAMD". In: *The Journal of Physical Chemistry Letters* 2.2 (2010), pp. 87–92.

[163] A. Dequidt, J. Devemy, and A. A. H. Pádua. "Thermalized Drude Oscillators with the LAMMPS Molecular Dynamics Simulator". In: *Journal of Chemical Information and Modeling* 56.1 (2015), pp. 260–268.

[164] J. Huang, J. A. Lemkul, P. K. Eastman, and A. D. MacKerell Jr. "Molecular Dynamics Simulations Using the Drude Polarizable Force Field on GPUs with OpenMM: Implementation, Validation, and Benchmarks". In: *Journal of Computational Chemistry* 39.21 (2018), pp. 1682–1689.

[165] G. Lamoureux, A. D. MacKerell Jr, and B. Roux. "A Simple Polarizable Model of Water Based on Classical Drude Oscillators". In: *The Journal of Chemical Physics* 119.10 (2003), pp. 5185–5197.

[166] G. Lamoureux, E. Harder, I. V. Vorobyov, B. Roux, and A. D. MacKerell Jr. "A Polarizable Model of Water for Molecular Dynamics Simulations of Biomolecules". In: *Chemical Physics Letters* 418.1-3 (2006), pp. 245–249.

[167] W. Yu, P. E. M. Lopes, B. Roux, and A. D. MacKerell Jr. "Six-Site Polarizable Model of Water Based on the Classical Drude Oscillator". In: *The Journal of Chemical Physics* 138.3 (2013), p. 34508.

[168] P. E. M. Lopes, J. Huang, J. Shim, Y. Luo, H. Li, B. Roux, and A. D. MacKerell Jr. "Polarizable Force Field for Peptides and Proteins Based on the Classical Drude Oscillator". In: *Journal of Chemical Theory and Computation* 9.12 (2013), pp. 5430–5449.

[169] B. T. Thole. "Molecular Polarizabilities Calculated With a Modified Dipole Interaction". In: *Chemical Physics* 59.3 (1981), pp. 341–350.

[170] J. Chowdhary, E. Harder, P. E. M. Lopes, L. Huang, A. D. MacKerell Jr, and B. Roux. "A Polarizable Force Field of Dipalmitoylphosphatidylcholine Based on the Classical Drude Model for Molecular Dynamics Simulations of Lipids". In: *The Journal of Physical Chemistry B* 117.31 (2013), pp. 9142–9160.

[171] H. Li, J. Chowdhary, L. Huang, X. He, A. D. MacKerell Jr, and B. Roux. "Drude Polarizable Force Field for Molecular Dynamics Simulations of Saturated and Unsaturated Zwitterionic Lipids". In: *Journal of Chemical Theory and Computation* 13.9 (2017), pp. 4535–4552.

[172] A. Savelyev and A. D. MacKerell Jr. "All-Atom Polarizable Force Field for DNA Based on the Classical Drude Oscillator Model". In: *Journal of Computational Chemistry* 35.16 (2014), pp. 1219–1239.

[173] J. A. Lemkul and A. D. MacKerell Jr. "Polarizable Force Field for DNA Based on the Classical Drude Oscillator: II. Microsecond Molecular Dynamics Simulations of Duplex DNA". In: *Journal of Chemical Theory and Computation* 13.5 (2017), pp. 2072–2085.

[174] J. A. Lemkul and A. D. MacKerell Jr. "Polarizable Force Field for RNA Based on the Classical Drude Oscillator". In: *Journal of Computational Chemistry* 39.32 (2018), pp. 2624–2646.

[175] D. S. Patel, X. He, and A. D. MacKerell. "Polarizable Empirical Force Field for Hexopyranose Monosaccharides Based on the Classical Drude Oscillator". In: *The Journal of Physical Chemistry B* 119.3 (2015), pp. 637–652.

[176] M. C. Small, A. H. Aytenfisu, F.-Y. Lin, X. He, and A. D. MacKerell. "Drude Polarizable Force Field for Aliphatic Ketones and Aldehydes, and Their Associated Acyclic Carbohydrates". In: *Journal of Computer-Aided Molecular Design* 31.4 (2017), pp. 349–363.

[177] F.-Y. Lin, P. E. M. Lopes, E. Harder, B. Roux, and A. D. MacKerell Jr. "Polarizable Force Field for Molecular Ions Based on the Classical Drude Oscillator". In: *Journal of Chemical Information and Modeling* 58.5 (2018), pp. 993–1004.

[178] S. Nosé. "A Unified Formulation of the Constant Temperature Molecular Dynamics Methods". In: *The Journal of Chemical Physics* 81.1 (1984), pp. 511–519.

[179] W. G. Hoover. "Canonical Dynamics: Equilibrium Phase-Space Distributions". In: *Physical Review A* 31.3 (1985), p. 1695.

[180] A. Brünger, C. L. Brooks III, and M. Karplus. "Stochastic Boundary Conditions for Molecular Dynamics Simulations of ST2 Water". In: *Chemical Physics Letters* 105.5 (1984), pp. 495–500.

[181] J. Applequist. "A Multipole Interaction Theory of Electric Polarization of Atomic and Molecular Assemblies". In: *The Journal of Chemical Physics* 83.2 (1985), pp. 809–826.

[182] A. Stone. *The Theory of Intermolecular Forces*. Oxford University Press, 2013.

[183] P. Ren and J. W. Ponder. "Consistent Treatment of Inter- and Intramolecular Polarization in Molecular Mechanics Calculations". In: *Journal of Computational Chemistry* 23.16 (2002), pp. 1497–1506.

[184] F. Lipparini, L. Lagardere, B. Stamm, E. Cances, M. Schnieders, P. Ren, Y. Maday, and J.-P. Piquemal. "Scalable Evaluation of Polarization Energy and Associated Forces in Polarizable Molecular Dynamics: I. Toward Massively Parallel Direct Space Computations". In: *Journal of Chemical Theory and Computation* 10.4 (2014), pp. 1638–1651.

[185] P. Ren and J. W. Ponder. "Polarizable Atomic Multipole Water Model for Molecular Mechanics Simulation". In: *The Journal of Physical Chemistry B* 107.24 (2003), pp. 5933–5947.

[186] M. L. Laury, L.-P. Wang, V. S. Pande, T. Head-Gordon, and J. W. Ponder. "Revised Parameters for the AMOEBA Polarizable Atomic Multipole Water Model". In: *The Journal of Physical Chemistry B* 119.29 (2015), pp. 9423–9437.

[187] A. Grossfield, P. Ren, and J. W. Ponder. "Ion Solvation Thermodynamics from Simulation with a Polarizable Force Field". In: *Journal of the American Chemical Society* 125.50 (2003), pp. 15671–15682.

[188] Z. Wang. "Polarizable Force Field Development, and Applications to Conformational Sampling and Free Energy Calculation". PhD thesis. Washington University in St. Louis, 2018.

[189] P. Ren, C. Wu, and J. W. Ponder. "Polarizable Atomic Multipole-Based Molecular Mechanics for Organic Molecules". In: *Journal of Chemical Theory and Computation* 7.10 (2011), pp. 3143–3161.

[190] Y. Shi, Z. Xia, J. Zhang, R. Best, C. Wu, J. W. Ponder, and P. Ren. "Polarizable Atomic Multipole-Based AMOEBA Force Field for Proteins". In: *Journal of Chemical Theory and Computation* 9.9 (2013), pp. 4046–4063.

[191] C. Zhang, C. Lu, Z. Jing, C. Wu, J.-P. Piquemal, J. W. Ponder, and P. Ren. "AMOEBA Polarizable Atomic Multipole Force Field for Nucleic Acids". In: *Journal of Chemical Theory and Computation* 14.4 (2018), pp. 2084–2108.

[192] H. Chu, X. Peng, Y. Li, Y. Zhang, H. Min, and G. Li. "Polarizable Atomic Multipole-Based Force Field for DOPC and POPE Membrane Lipids". In: *Molecular Physics* 116.7-8 (2018), pp. 1037–1050.

[193] H. Chu, X. Peng, Y. Li, Y. Zhang, and G. Li. "A Polarizable Atomic Multipole-Based Force Field for Molecular Dynamics Simulations of Anionic Lipids". In: *Molecules* 23.1 (2018), p. 77.

[194] J. A. Rackers, Z. Wang, C. Lu, M. L. Laury, L. Lagardere, M. J. Schnieders, J.-P. Piquemal, P. Ren, and J. W. Ponder. "Tinker 8: Software Tools for Molecular Design". In: *Journal of Chemical Theory and Computation* 14.10 (2018), pp. 5273–5289.

[195] D. A. Case, V. Babin, J. Berryman, R. M. Betz, Q. Cai, D. S. Cerutti, T. E. Cheatham Iii, T. A. Darden, R. E. Duke, H. Gohlke, et al. *Amber 14*. University of California, 2014.

[196] P. Eastman, J. Swails, J. D. Chodera, R. T. McGibbon, Y. Zhao, K. A. Beauchamp, L.-P. Wang, A. C. Simmonett, M. P. Harrigan, C. D. Stern, et al. "OpenMM 7: Rapid Development of High Performance Algorithms for Molecular Dynamics". In: *PLOS Computational Biology* 13.7 (2017), e1005659.

[197] L. Lagardère, L.-H. Jolly, F. Lipparini, F. Aviat, B. Stamm, Z. F. Jing, M. Harger, H. Torabifard, G. A. Cisneros, M. J. Schnieders, et al. "Tinker-HP: A Massively Parallel Molecular Dynamics Package for Multiscale Simulations of Large Complex Systems With Advanced Point Dipole Polarizable Force Fields". In: *Chemical Science* 9.4 (2018), pp. 956–972.

[198] E. B. Wilson, J. C. Decius, and P. C. Cross. *Molecular Vibrations: The Theory of Infrared and Raman Vibrational Spectra*. Courier Corporation, 1980.

[199] T. A. Halgren. "The Representation of van der Waals (vdW) Interactions in Molecular Mechanics Force Fields: Potential Form, Combination Rules, and vdW Parameters". In: *Journal of the American Chemical Society* 114.20 (1992), pp. 7827–7843.

[200] C. J. Burnham, J. Li, S. S. Xantheas, and M. Leslie. "The Parametrization of a Thole-Type All-Atom Polarizable Water Model from First Principles and Its Application to the Study of Water Clusters ($n = 2-21$) and the Phonon Spectrum of Ice Ih". In: *The Journal of Chemical Physics* 110.9 (1999), pp. 4566–4581.

[201] L.-P. Wang, T. J. Martinez, and V. S. Pande. "Building Force Fields: An Automatic, Systematic, and Reproducible Approach". In: *The Journal of Physical Chemistry Letters* 5.11 (2014), pp. 1885–1891.

[202] A. K. Das, O. N. Demerdash, and T. Head-Gordon. "Improvements to the AMOEBA Force Field by Introducing Anisotropic Atomic Polarizability of the Water Molecule". In: *Journal of Chemical Theory and Computation* 14.12 (2018), pp. 6722–6733.

[203] R. Qi, L.-P. Wang, Q. Wang, V. S. Pande, and P. Ren. "United Polarizable Multipole Water Model for Molecular Mechanics Simulation". In: *The Journal of Chemical Physics* 143.1 (2015), p. 14504.

[204] L.-P. Wang, T. Head-Gordon, J. W. Ponder, P. Ren, J. D. Chodera, P. K. Eastman, T. J. Martinez, and V. S. Pande. "Systematic Improvement of a Classical Molecular Model of Water". In: *The Journal of Physical Chemistry B* 117.34 (2013), pp. 9956–9972.

[205] P. Pulay. "Convergence Acceleration of Iterative Sequences. The Case of SCF Iteration". In: *Chemical Physics Letters* 73.2 (1980), pp. 393–398.

[206] A. C. Simmonett, F. C. Pickard IV, Y. Shao, T. E. Cheatham III, and B. R. Brooks. "Efficient Treatment of Induced Dipoles". In: *The Journal of Chemical Physics* 143.7 (2015), p. 74115.

[207] M. Tuckerman, B. J. Berne, and G. J. Martyna. "Reversible Multiple Time Scale Molecular Dynamics". In: *The Journal of Chemical Physics* 97.3 (1992), pp. 1990–2001.

[208] G. M. Torrie and J. P. Valleau. "Monte Carlo Free Energy Estimates Using Non-Boltzmann Sampling: Application to the Sub-Critical Lennard-Jones Fluid". In: *Chemical Physics Letters* 28.4 (1974), pp. 578–581.

[209] S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kollman. "The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules. I. The Method". In: *Journal of Computational Chemistry* 13.8 (1992), pp. 1011–1021.

[210] J. S. Hub, B. L. De Groot, and D. Van Der Spoel. "g_wham A Free Weighted Histogram Analysis Implementation Including Robust Error and Autocorrelation Estimates". In: *Journal of Chemical Theory and Computation* 6.12 (2010), pp. 3713–3720.

[211] A. Grossfield. *WHAM: The Weighted Histogram Analysis Method, Version 2.0.9.* 2013. URL: `http://membrane.urmc.rochester.edu/wordpress/?page%7B%5C_%7Did=126`.

[212] B. Honig and A. Nicholls. "Classical Electrostatics in Biology and Chemistry". In: *Science* 268.5214 (1995), pp. 1144–1149.

[213] M. E. Davis and J. A. McCammon. "Electrostatics in Biomolecular Structure and Dynamics". In: *Chemical Reviews* 90.3 (1990), pp. 509–521.

[214] M. Holst and F. Saied. "Multigrid Solution of the Poisson-Boltzmann Equation". In: *Journal of Computational Chemistry* 14.1 (1993), pp. 105–113.

[215] M. J. Holst and F. Saied. "Numerical Solution of the Nonlinear Poisson–Boltzmann Equation: Developing More Robust and Efficient Methods". In: *Journal of Computational Chemistry* 16.3 (1995), pp. 337–364.

[216] S. T. Kottmann. "Harmonic Averaging of Smooth Permittivity Functions in Finite-Difference Poisson-Boltzmann Electrostatics". In: *Theoretical Chemistry Accounts* 119.5-6 (2008), pp. 421–427.

[217] E. Huckel and P. Debye. "Zur Theorie der Elektrolyte. I. Gefrierpunktserniedrigung und verwandte Erscheinungen". In: *Physikalische Zeitschrift* 24 (1923), pp. 185–206.

[218] N. A. Baker, D. Sept, S. Joseph, M. J. Holst, and J. A. McCammon. "Electrostatics of Nanosystems: Application to Microtubules and the Ribosome". In: *Proceedings of the National Academy of Sciences* 98.18 (2001), pp. 10037–10041.

[219] E. Jurrus, D. Engel, K. Star, K. Monson, J. Brandi, L. E. Felberg, D. H. Brookes, L. Wilson, J. Chen, K. Liles, et al. "Improvements to the APBS Biomolecular Solvation Software Suite". In: *Protein Science* 27.1 (2018), pp. 112–128.

[220] M. Krone, B. Kozlíková, N. Lindow, M. Baaden, D. Baum, J. Parulek, H. -C. Hege, and I. Viola. "Visual Analysis of Biomolecular Cavities: State of the Art". In: *Computer Graphics Forum* 35.3 (2016), pp. 527–551.

[221] O. S. Smart, J. M. Goodfellow, and B. Wallace. "The Pore Dimensions of Gramicidin A". In: *Biophysical Journal* 65.6 (1993), pp. 2455–2460.

[222] O. S. Smart, J. G. Neduvelil, X. Wang, B. A. Wallace, and M. S. P. Sansom. "HOLE: A Program for the Analysis of the Pore Dimensions of Ion Channel Structural Models". In: *Journal of Molecular Graphics* 14.6 (1996), pp. 354–360.

[223] M. Petrek, M. Otyepka, P. Banás, P. Kosinová, J. Koca, and J. Damborský. "CAVER: A New Tool to Explore Routes From Protein Clefts, Pockets and Cavities". In: *BMC Bioinformatics* 7 (2006), p. 316.

[224] E. Chovancova, A. Pavelka, P. Benes, O. Strnad, J. Brezovsky, B. Kozlikova, A. Gora, V. Sustr, M. Klvana, P. Medek, L. Biedermannova, J. Sochor, and J. Damborsky. "CAVER 3.0: A Tool for the Analysis of Transport Pathways in Dynamic Protein Structures". In: *PLOS Computational Biology* 8.10 (2012), e1002708.

[225] S. Basak, Y. Gicheru, S. Rao, M. S. P. Sansom, and S. Chakrapani. "Cryo-EM Reveals Two Distinct Serotonin-Bound Conformations of Full-Length 5-$HT_{3A}$ Receptor". In: *Nature* 563.7730 (2018), pp. 270–274.

[226] S. Yuan, S. Filipek, and H. Vogel. "A Gating Mechanism of the Serotonin 5-$HT_3$ Receptor". In: *Structure* 24.5 (2016), pp. 816–825.

[227] P. J. Stansfeld and M. S. P. Sansom. "From Coarse Grained to Atomistic: A Serial Multiscale Approach to Membrane Protein Simulations". In: *Journal of Chemical Theory and Computation* 7.4 (2011), pp. 1157–1166.

[228] J. Huang and A. D. MacKerell Jr. "CHARMM36 All-Atom Additive Protein Force Field: Validation Based on Comparison to NMR Data". In: *Journal of Computational Chemistry* 34.25 (2013), pp. 2135–2145.

[229] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Science & Business Media, 2013.

[230] S. J. Sheather and M. C. Jones. "A Reliable Data-Based Bandwidth Selection Method for Kernel Density Estimation". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 53.3 (1991), pp. 683–690.

[231] J. A. Nelder and R. Mead. "A Simplex Method for Function Minimization". In: *The Computer Journal* 7.4 (1965), pp. 308–313.

[232] A. Gharpure, J. Teng, Y. Zhuang, C. M. Noviello, R. M. Walsh Jr, R. Cabuco, R. J. Howard, N. T. Zaveri, E. Lindahl, and R. E. Hibbs. "Agonist Selectivity and Ion Permeation in the $\alpha 3\beta 4$ Ganglionic Nicotinic Receptor". In: *Neuron* (2019).

[233] R. P. Brent. "An Algorithm With Guaranteed Convergence for Finding a Zero of a Function". In: *The Computer Journal* 14.4 (1971), pp. 422–425.

[234] A. P. Willard and D. Chandler. "Instantaneous Liquid Interfaces". In: *The Journal of Physical Chemistry B* 114.5 (2010), pp. 1954–1958.

[235] J. Kyte and R. F. Doolittle. "A Simple Method for Displaying the Hydropathic Character of a Protein". In: *Journal of Molecular Biology* 157.1 (1982), pp. 105–132.

[236] O. D. Monera, T. J. Sereda, N. E. Zhou, C. M. Kay, and R. S. Hodges. "Relationship of Sidechain Hydrophobicity and $\alpha$-Helical Propensity on the Stability of the Single-Stranded Amphipathic $\alpha$-Helix". In: *Journal of Peptide Science* 1.5 (1995), pp. 319–329.

[237] W. C. Wimley, T. P. Creamer, and S. H. White. "Solvation Energies of Amino Acid Side Chains and Backbone in a Family of Host-Guest Pentapeptides". In: *Biochemistry* 35.16 (1996), pp. 5109–5124.

[238] W. C. Wimley and S. H. White. "Experimentally Determined Hydrophobicity Scale for Proteins at Membrane Interfaces". In: *Nature Structural Biology* 3.10 (1996), pp. 842–848.

[239] C. P. Moon and K. G. Fleming. "Side-Chain Hydrophobicity Scale Derived from Transmembrane Protein Folding into Lipid Bilayers". In: *Proceedings of the National Academy of Sciences* 108.25 (2011), pp. 10174–10177.

[240] T. Hessa, H. Kim, K. Bihlmaier, C. Lundin, J. Boekel, H. Andersson, I. Nilsson, S. H. White, and G. von Heijne. "Recognition of Transmembrane Helices by the Endoplasmic Reticulum Translocon". In: *Nature* 433.7024 (2005), pp. 377–381.

[241] C. Zhu, Y. Gao, H. Li, S. Meng, L. Li, J. S. Francisco, and X. C. Zeng. "Characterizing Hydrophobicity of Amino Acid Side Chains in a Protein Environment Via Measuring Contact Angle of a Water Nanodroplet on Planar Peptide Network". In: *Proceedings of the National Academy of Sciences* 113.46 (2016), pp. 12946–12951.

[242] G. Vriend. "WHAT IF: A Molecular Modeling and Drug Design Program". In: *Journal of Molecular Graphics* 8.1 (1990), pp. 52–56.

[243] M. S. Floater and T. Surazhsky. "Parameterization for Curve Interpolation". In: *Studies in Computational Mathematics*. 2006, pp. 39–54.

[244] E. Anderson, Z. Bai, J. Dongarra, A. Greenbaum, A. McKenney, J. Du Croz, S. Hammarling, J. Demmel, C. Bischof, and D. Sorensen. "LAPACK: A Portable Linear Algebra Library for High-Performance Computers". In: *Proceedings of the 1990 ACM/IEEE Conference on Supercomputing*. IEEE Computer Society Press. 1990, pp. 2–11.

[245] H. Wang, J. Kearney, and K. Atkinson. "Arc-Length Parameterized Spline Curves for Real-Time Simulation". In: *Proceedings of the 5th International Conference on Curves and Surfaces*. 2002, pp. 387–396.

[246] G. E. Alefeld, F. A. Potra, and Y. Shi. "Algorithm 748: Enclosing Zeros of Continuous Functions". In: *ACM Transactions on Mathematical Software* 21.3 (1995), pp. 327–344.

[247] G. Klesse. *CHAP Github Repository*. 2019. URL: `www.github.com/channotation/chap`.

[248] W. Humphrey, A. Dalke, and K. Schulten. "VMD: Visual Molecular Dynamics". In: *Journal of Molecular Graphics* 14.1 (1996), pp. 33–38.

[249] W. L. DeLano. "PyMOL: An Open-Source Molecular Graphics Tool". In: *CCP4 Newsletter On Protein Crystallography* 40 (2002), pp. 82–92.

[250] A. Morin, B. Eisenbraun, J. Key, P. C. Sanschagrin, M. A. Timony, M. Ottaviano, and P. Sliz. "Cutting Edge: Collaboration Gets the Most out of Software". In: *elife* 2 (2013), e01456.

[251] SBGrid Consortium. *SBGrid Page for CHAP*. 2019. URL: `https://sbgrid.org/software/titles/chap`.

[252] G. L. Isom, N. Coudray, M. R. MacRae, C. T. McManus, D. C. Ekiert, and G. Bhabha. "Structure of LetB Reveals a Tunnel for Lipid Transport Across the Bacterial Envelope". In: *bioRxiv* (2019), p. 748145.

[253] M. Igaev, C. Kutzner, L. V. Bock, A. C. Vaiana, and H. Grubmüller. "Automated Cryo-EM Structure Refinement Using Correlation-Driven Molecular Dynamics". In: *eLife* 8 (2019), e43542.

[254] M. Calvelo, J. R. Granja, and R. Garcia-Fandino. "Competitive double-switched self-assembled cyclic peptide nanotubes: a dual internal and external control". In: *Physical Chemistry Chemical Physics* 21 (2019), pp. 20750–20756.

[255] G. Hassaine, C. Deluz, L. Grasso, R. Wyss, M. B. Tol, R. Hovius, A. Graff, H. Stahlberg, T. Tomizaki, A. Desmyter, et al. "X-Ray Structure of the Mouse Serotonin 5-$HT_3$ Receptor". In: *Nature* 512.7514 (2014), pp. 276–281.

[256] S. Basak, Y. Gicheru, A. Samanta, S. K. Molugu, W. Huang, M. la de Fuente, T. Hughes, D. J. Taylor, M. T. Nieman, V. Moiseenkova-Bell, et al. "Cryo-EM Structure of 5-$HT_{3A}$ Receptor in Its Resting Conformation". In: *Nature Communications* 9.1 (2018), p. 514.

[257] X. Huang, H. Chen, K. Michelsen, S. Schneider, and P. L. Shaffer. "Crystal Structure of Human Glycine Receptor-$\alpha$3 Bound to Antagonist Strychnine". In: *Nature* 526.7572 (2015), pp. 277–280.

[258] X. Huang, H. Chen, and P. L. Shaffer. "Crystal Structures of Human GlyR$\alpha$3 Bound to Ivermectin". In: *Structure* 25.6 (2017), pp. 945–950.

[259] J. Du, W. Lü, S. Wu, Y. Cheng, and E. Gouaux. "Glycine Receptor Mechanism Elucidated by Electron Cryo-Microscopy". In: *Nature* 526.7572 (2015), pp. 224–229.

[260] G. Gonzalez-Gutierrez, Y. Wang, G. D. Cymes, E. Tajkhorshid, and C. Grosman. "Chasing the Open-State Structure of Pentameric Ligand-Gated Ion Channels". In: *The Journal of General Physiology* 149.12 (2017), pp. 1119–1138.

[261] A. H. Cerdan, N. É. Martin, and M. Cecchini. "An Ion-Permeable State of the Glycine Receptor Captured by Molecular Dynamics". In: *Structure* 26.11 (2018), pp. 1555–1562.

[262] M. A. Dämgen and P. C. Biggin. "A Refined Open State of the Glycine Receptor Obtained Via Molecular Dynamics Simulations". In: *bioRxiv* (2019), p. 668830.

[263] L. Polovinkin, G. Hassaine, J. Perot, E. Neumann, A. A. Jensen, S. N. Lefebvre, P.-J. Corringer, J. Neyton, C. Chipot, F. Dehez, et al. "Conformational Transitions of the Serotonin 5-$HT_3$ Receptor". In: *Nature* 563.7730 (2018), pp. 275–279.

[264] X. Huang, P. L. Shaffer, S. Ayube, H. Bregman, H. Chen, S. G. Lehto, J. A. Luther, D. J. Matson, S. I. McDonough, K. Michelsen, et al. "Crystal Structures of Human Glycine Receptor $\alpha 3$ Bound to a Novel Class of Analgesic Potentiators". In: *Nature Structural & Molecular Biology* 24.2 (2017), pp. 108–113.

[265] W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives. "Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids". In: *Journal of the American Chemical Society* 118.45 (1996), pp. 11225–11236.

[266] A. J. Thompson and S. R Lummis. "5-$HT_3$ Receptors". In: *Current Pharmaceutical Design* 12.28 (2006), pp. 3615–3630.

[267] J. Walstab, G. Rappold, and B. Niesler. "5-$HT_3$ Receptors: Role in Disease and Target of Drugs". In: *Pharmacology & Therapeutics* 128.1 (2010), pp. 146–169.

[268] S. C. Lummis. "5-$HT_3$ Receptors". In: *Journal of Biological Chemistry* 287.48 (2012), pp. 40239–40245.

[269] T. K. Machu. "Therapeutics of 5-$HT_3$ Receptor Antagonists: Current Uses and Future Directions". In: *Pharmacology & Therapeutics* 130.3 (2011), pp. 338–347.

[270] M. Fayyaz and J. M. Lackner. "Serotonin Receptor Modulators in the Treatment of Irritable Bowel Syndrome". In: *Therapeutics and Clinical Risk Management* 4.1 (2008), pp. 41–48.

[271] M. S. Aapro. "5-$HT_3$ Receptor Antagonists". In: *Drugs* 42.4 (1991), pp. 551–568.

[272] J. Mähler and I. Persson. "A Study of the Hydration of the Alkali Metal Ions in Aqueous Solution". In: *Inorganic Chemistry* 51.1 (2011), pp. 425–438.

[273] A. des Georges, O. B. Clarke, R. Zalk, Q. Yuan, K. J. Condon, R. A. Grassucci, W. A. Hendrickson, A. R. Marks, and J. Frank. "Structural Basis for Gating and Activation of RyR1". In: *Cell* 167.1 (2016), pp. 145–157.

[274] S. Andreev, D. Reichman, and G. Hummer. "Effect of Flexibility on Hydrophobic Behavior of Nanotube Water Channels". In: *The Journal of Chemical Physics* 123.19 (2005), p. 194502.

[275] Y. E. Altabet and P. G. Debenedetti. "The Role of Material Flexibility on the Drying Transition of Water Between Hydrophobic Objects: A Thermodynamic Analysis". In: *The Journal of Chemical Physics* 141.18 (2014), p. 18C531.

[276] Y. E. Altabet, A. Haji-Akbari, and P. G. Debenedetti. "Effect of Material Flexibility on the Thermodynamics and Kinetics of Hydrophobically Induced Evaporation of Water". In: *Proceedings of the National Academy of Sciences* 114.13 (2017), E2548–E2555.

[277] J. W. Lynch. "Molecular Structure and Function of the Glycine Receptor Chloride Channel". In: *Physiological Reviews* 84.4 (2004), pp. 1051–1095.

[278] J. W. Lynch. "Native Glycine Receptor Subtypes and Their Physiological Roles". In: *Neuropharmacology* 56.1 (2009), pp. 303–309.

[279] C. F. Burgos, G. E. Yévenes, and L. G. Aguayo. "Structure and Pharmacologic Modulation of Inhibitory Glycine Receptors". In: *Molecular Pharmacology* 90.3 (2016), pp. 318–325.

[280] J. W. Lynch and R. J. Callister. "Glycine Receptors: A New Therapeutic Target in Pain Pathways". In: *Current Opinion in Investigational Drugs* 7.1 (2006), pp. 48–53.

[281] A. Bode and J. W. Lynch. "The Impact of Human Hyperekplexia Mutations on Glycine Receptor Structure and Function". In: *Molecular Brain* 7.1 (2014), p. 2.

[282] A. J. Thompson, H. A. Lester, and S. C. Lummis. "The Structural Basis of Function in Cys-Loop Receptors". In: *Quarterly Reviews of Biophysics* 43.4 (2010), pp. 449–499.

[283] P. L. Silvestrelli and M. Parrinello. "Water Molecule Dipole in the Gas and in the Liquid Phase". In: *Physical Review Letters* 82.16 (1999), pp. 3308–3311.

[284] T. R. Dyke and J. S. Muenter. "Electric Dipole Moments of Low $J$ States of $H_2O$ and $D_2O$". In: *The Journal of Chemical Physics* 59.6 (1973), pp. 3125–3127.

[285] V. A. Ngo, J. K. Fanning, and S. Y. Noskov. "Comparative Analysis of Protein Hydration from MD Simulations With Additive and Polarizable Force Fields". In: *Advanced Theory and Simulations* 2.2 (2019), p. 1800106.

[286] N. B. Vargaftik, B. N. Volkov, and L. D. Voljak. "International Tables of the Surface Tension of Water". In: *Journal of Physical and Chemical Reference Data* 12.3 (1983), pp. 817–820.

[287] R. Mills. "Self-Diffusion in Normal and Heavy Water in the Range 1–45°". In: *The Journal of Physical Chemistry* 77.5 (1973), pp. 685–688.

[288] J. Pátek, J. Hrubý, J. Klomfar, M. Součková, and A. H. Harvey. "Reference Correlations for Thermophysical Properties of Liquid Water at 0.1 MPa". In: *Journal of Physical and Chemical Reference Data* 38.1 (2009), pp. 21–29.

[289] C. Vega, J. L. F. Abascal, M. M. Conde, and J. L. Aragones. "What Ice Can Teach Us About Water Interactions: A Critical Comparison of the Performance of Different Water Models". In: *Faraday Discussions* 141 (2009), pp. 251–276.

[290] F. Chen and P. E. Smith. "Simulated Surface Tensions of Common Water Models". In: *The Journal of Chemical Physics* 126.22 (2007), p. 221101.

[291] T. A. Halgren and W. Damm. "Polarizable Force Fields". In: *Current Opinion in Structural Biology* 11.2 (2001), pp. 236–242.

[292] A. D. MacKerell Jr. "Empirical Force Fields for Biological Macromolecules: Overview and Issues". In: *Journal of Computational Chemistry* 25.13 (2004), pp. 1584–1604.

[293] O. Demerdash, L.-P. Wang, and T. Head-Gordon. "Advanced Models for Water Simulations". In: *Wiley Interdisciplinary Reviews: Computational Molecular Science* 8.1 (2018), e1355.

[294] Z. Jing, C. Liu, S. Y. Cheng, R. Qi, B. D. Walker, J.-P. Piquemal, and P. Ren. "Polarizable Force Fields for Biomolecular Simulations: Recent Advances and Applications". In: *Annual Review of Biophysics* 48 (2019).

[295] P. Cieplak, J. Caldwell, and P. Kollman. "Molecular Mechanical Models for Organic and Biological Systems Going Beyond the Atom Centered Two Body Additive Approximation: Aqueous Solution Free Energies of Methanol and N-Methyl Acetamide, Nucleic Acid Base, and Amide Hydrogen Bonding and Chloroform/Water Partition Coefficients of the Nucleic Acid Bases". In: *Journal of Computational Chemistry* 22.10 (2001), pp. 1048–1057.

[296] B. A. Bauer and S. Patel. "Recent Applications and Developments of Charge Equilibration Force Fields for Modeling Dynamical Charges in Classical Molecular Dynamics Simulations". In: *Theoretical Chemistry Accounts* 131.3 (2012), p. 1153.

[297] P. Ren and J. W. Ponder. "Temperature and Pressure Dependence of the AMOEBA Water Model". In: *The Journal of Physical Chemistry B* 108.35 (2004), pp. 13427–13437.

[298] D. M. Chipman. "Water from Ambient to Supercritical Conditions with the AMOEBA Model". In: *The Journal of Physical Chemistry B* 117.17 (2013), pp. 5148–5155.

[299] M. R. Stukan, A. Asmadi, and W. Abdallah. "Bulk Properties of SWM4-NDP Water Model at Elevated Temperature and Pressure". In: *Journal of Molecular Liquids* 180 (2013), pp. 65–69.

[300] H. Yu, T. W. Whitfield, E. Harder, G. Lamoureux, I. Vorobyov, V. M. Anisimov, A. D. MacKerell Jr, and B. Roux. "Simulating Monovalent and Divalent Ions in Aqueous Solution Using a Drude Polarizable Force Field". In: *Journal of Chemical Theory and Computation* 6.3 (2010), pp. 774–786.

[301] J. L. F. Abascal, E. Sanz, and C. Vega. "Triple Points and Coexistence Properties of the Dense Phases of Water Calculated Using Computer Simulation". In: *Physical Chemistry Chemical Physics* 11.3 (2009), pp. 556–562.

[302] S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman, and A. H. De Vries. "The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations". In: *The Journal of Physical Chemistry B* 111.27 (2007), pp. 7812–7824.

[303] J. B. Klauda, V. Monje, T. Kim, and W. Im. "Improving the CHARMM Force Field for Polyunsaturated Fatty Acid Chains". In: *The Journal of Physical Chemistry B* 116.31 (2012), pp. 9424–9431.

[304] S. Jo, T. Kim, V. G. Iyer, and W. Im. "CHARMM-GUI: A Web-Based Graphical User Interface for CHARMM". In: *Journal of Computational Chemistry* 29.11 (2008), pp. 1859–1865.

[305] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, and K. Schulten. "Scalable Molecular Dynamics With NAMD". In: *Journal of Computational Chemistry* 26.16 (2005), pp. 1781–1802.

[306] C. Sagui, L. G. Pedersen, and T. A. Darden. "Towards an Accurate Representation of Electrostatics in Classical Force Fields: Efficient Implementation of Multipolar Interactions in Biomolecular Simulations". In: *The Journal of Chemical Physics* 120.1 (2004), pp. 73–87.

[307] R. J. Gowers, M. Linke, J. Barnoud, T. J. E. Reddy, M. N. Melo, S. L. Seyler, J. Domański, D. L. Dotson, S. Buchoux, I. M. Kenney, et al. "MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations". In: *Proceedings of the 15th Python in Science Conference*. Vol. 98. SciPy Austin, TX. 2016.

[308] A. Bankura, V. Carnevale, and M. L. Klein. "Hydration Structure of Salt Solutions from Ab Initio Molecular Dynamics". In: *The Journal of Chemical Physics* 138.1 (2013), p. 14501.

[309] C. N. Rowley and B. Roux. "The Solvation Structure of $Na^+$ and $K^+$ in Liquid Water Determined From High Level Ab Initio Molecular Dynamics Simulations". In: *Journal of Chemical Theory and Computation* 8.10 (2012), pp. 3526–3535.

[310] S. Varma and S. B. Rempe. "Coordination Numbers of Alkali Metal Ions in Aqueous Solutions". In: *Biophysical chemistry* 124.3 (2006), pp. 192–199.

[311] D. H. Powell, G. W. Neilson, and J. E. Enderby. "The Structure of $Cl^-$ Aqueous Solution: An Experimental Determination of $g_{\mathrm{ClH}}(r)$ and $g_{\mathrm{ClO}}(r)$". In: *Journal of Physics: Condensed Matter* 5.32 (1993), pp. 5723–5730.

[312] S. Cummings, J. E. Enderby, G. W. Neilson, J. R. Newsome, R. A. Howe, W. S. Howells, and A. K. Soper. "Chloride Ions in Aqueous Solutions". In: *Nature* 287.5784 (1980), pp. 714–716.

[313] R. D'Auria and D. J. Tobias. "Relation Between Surface Tension and Ion Adsorption at the Air-Water Interface: A Molecular Dynamics Simulation Study". In: *The Journal of Physical Chemistry A* 113.26 (2009), pp. 7286–7293.

[314] L.-P. Wang, K. A. McKiernan, J. Gomes, K. A. Beauchamp, T. Head-Gordon, J. E. Rice, W. C. Swope, T. J. Martinez, and V. S. Pande. "Building a More Predictive Protein Force Field: A Systematic and Reproducible Route to AMBER-FB15". In: *The Journal of Physical Chemistry B* 121.16 (2017), pp. 4023–4039.

[315] K. T. Debiec, D. S. Cerutti, L. R. Baker, A. M. Gronenborn, D. A. Case, and L. T. Chong. "Further Along the Road Less Traveled: AMBER ff15ipq, an Original Protein Force Field Built on a Self-Consistent Physical Model". In: *Journal of Chemical Theory and Computation* 12.8 (2016), pp. 3926–3947.

[316] K. Takemura and A. Kitao. "Water Model Tuning for Improved Reproduction of Rotational Diffusion and NMR Spectral Density". In: *The Journal of Physical Chemistry B* 116.22 (2012), pp. 6279–6287.

[317] D. L. Mobley, K. L. Wymer, N. M. Lim, and J. P. Guthrie. "Blind Prediction of Solvation Free Energies from the SAMPL4 Challenge". In: *Journal of Computer-Aided Molecular Design* 28.3 (2014), pp. 135–150.

[318] R. T. Bradshaw and J. W. Essex. "Evaluating Parametrization Protocols for Hydration Free Energy Calculations with the AMOEBA Polarizable Force Field". In: *Journal of Chemical Theory and Computation* 12.8 (2016), pp. 3871–3883.

[319] C. D. Daub, D. Bratko, K. Leung, and A. Luzar. "Electrowetting at the Nanoscale". In: *The Journal of Physical Chemistry C* 111.2 (2007), pp. 505–509.

[320] D. Bratko, C. D. Daub, K. Leung, and A. Luzar. "Effect of Field Direction on Electrowetting in a Nanopore". In: *Journal of the American Chemical Society* 129.9 (2007), pp. 2504–2510.

[321] M. Sotomayor, V. Vásquez, E. Perozo, and K. Schulten. "Ion Conduction Through MscS as Determined by Electrophysiology and Simulation". In: *Biophysical Journal* 92.3 (2007), pp. 886–902.

[322] U. Rant. "Nanopores: Water Flow at the Flip of a Switch". In: *Nature Nanotechnology* 6.12 (2011), pp. 759–760.

[323] S. N. Smirnov, I. V. Vlassiouk, and N. V. Lavrik. "Voltage-Gated Hydrophobic Nanopores". In: *ACS Nano* 5.9 (2011), pp. 7453–7461.

[324] K. Xiao, Y. Zhou, X.-Y. Kong, G. Xie, P. Li, Z. Zhang, L. Wen, and L. Jiang. "Electrostatic-Charge- and Electric-Field-Induced Smart Gating for Water Transportation". In: *ACS Nano* 10.10 (2016), pp. 9703–9709.

[325] J. Gumbart, F. Khalili-Araghi, M. Sotomayor, and B. Roux. "Constant Electric Field Simulations of the Membrane Potential Illustrated with Simple Systems". In: *Biochimica et Biophysica Acta — Biomembranes* 1818.2 (2012), pp. 294–302.

[326] M. Tarek. "Membrane Electroporation: A Molecular Dynamics Simulation". In: *Biophysical Journal* 88.6 (2005), pp. 4045–4053.

[327] R. A. Böckmann, B. L. De Groot, S. Kakorin, E. Neumann, and H. Grubmüller. "Kinetics, Statistics, and Energetics of Lipid Membrane Electroporation Studied by Molecular Dynamics Simulations". In: *Biophysical Journal* 95.4 (2008), pp. 1837–1850.

[328] L. Delemotte and M. Tarek. "Molecular Dynamics Simulations of Lipid Membrane Electroporation". In: *The Journal of Membrane Biology* 245.9 (2012), pp. 531–543.

[329] M. K. Gilson and B. H. Honig. "Calculation of Electrostatic Potentials in an Enzyme Active Site". In: *Nature* 330.6143 (1987), pp. 84–86.

[330] A. A. Rashin and B. Honig. "Reevaluation of the Born Model of Ion Hydration". In: *The Journal of Physical Chemistry* 89.26 (1985), pp. 5588–5593.

[331] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria, 2018. URL: `https://www.R-project.org/`.

[332] A. Gelman and J. Hill. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press, 2006.

[333] B. Carpenter, A. Gelman, M. D. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. Brubaker, J. Guo, P. Li, and A. Riddell. "Stan: A Probabilistic Programming Language". In: *Journal of Statistical Software* 76.1 (2017).

[334] P.-C. Bürkner et al. "brms: An R Package for Bayesian Multilevel Models using Stan". In: *Journal of Statistical Software* 80.1 (2017), pp. 1–28.

[335] R. M. Neal et al. "MCMC Using Hamiltonian Dynamics". In: *Handbook of Markov Chain Monte Carlo* 2.11 (2011), p. 2.

[336] M. Betancourt. *A Conceptual Introduction to Hamiltonian Monte Carlo*. 2017. arXiv: 1701.02434.

[337] M. D. Hoffman and A. Gelman. "The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo". In: *Journal of Machine Learning Research* 15.1 (2014), pp. 1593–1623.

[338] G. Wilkinson and C. Rogers. "Symbolic Description of Factorial Models for Analysis of Variance". In: *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 22.3 (1973), pp. 392–399.

[339] A. J. Thompson and S. C. Lummis. "A Single Ring of Charged Amino Acids at One End of The Pore Can Control Ion Selectivity in the 5-$HT_3$ Receptor". In: *British Journal of Pharmacology* 140.2 (2003), pp. 359–365.

[340] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, and T. E. Ferrin. "UCSF Chimera – A Visualization System for Exploratory Research and Analysis". In: *Journal of Computational Chemistry* 25.13 (2004), pp. 1605–1612.

[341] J. Lee, X. Cheng, J. M. Swails, M. S. Yeom, P. K. Eastman, J. A. Lemkul, S. Wei, J. Buckner, J. C. Jeong, Y. Qi, et al. "CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field". In: *Journal of Chemical Theory and Computation* 12.1 (2015), pp. 405–413.

[342] J.-P. Ebejer, S. Fulle, G. M. Morris, and P. W. Finn. "The Emerging Role of Cloud Computing in Molecular Modelling". In: *Journal of Molecular Graphics and Modelling* 44 (2013), pp. 177–187.

[343] N. D. Hine, P. D. Haynes, A. A. Mostofi, C.-K. Skylaris, and M. C. Payne. "Linear-Scaling Density-Functional Theory With Tens of Thousands of Atoms: Expanding the Scope and Scale of Calculations with ONETEP". In: *Computer Physics Communications* 180.7 (2009), pp. 1041–1053.

[344] A. P. Jones, J. Crain, V. P. Sokhan, T. W. Whitfield, and G. J. Martyna. "Quantum Drude Oscillator Model of Atoms and Molecules: Many-Body Polarization and Dispersion Interactions for Atomistic Simulation". In: *Physical Review B* 87.14 (2013), p. 144103.

[345] A. Jones, J. Crain, F. Cipcigan, V. Sokhan, M. Modani, and G. Martyna. "Electronically Coarse-Grained Molecular Dynamics Using Quantum Drude Oscillators". In: *Molecular Physics* 111.22-23 (2013), pp. 3465–3477.

[346] A. Jones, F. Cipcigan, V. Sokhan, J. Crain, and G. Martyna. "Electronically Coarse-Grained Model for Water". In: *Physical Review Letters* 110.22 (2013), p. 227801.

[347] V. P. Sokhan, A. P. Jones, F. S. Cipcigan, J. Crain, and G. J. Martyna. "Signature Properties of Water: Their Molecular Electronic Origins". In: *Proceedings of the National Academy of Sciences* 112.20 (2015), pp. 6341–6346.