

COMPUTATIONAL STUDIES OF SMALL MOLECULE  
PERMEATION ACROSS MEMBRANE CHANNELS

D I S S E R T A T I O N  
for the award of the degree  
“Doctor rerum naturalium”  
of the Georg-August-Universität Göttingen

within the doctoral program  
Physics of Biological and Complex Systems (PBCS)  
of the Georg-August University School of Science (GAUSS)

Submitted by  
Igor Ariz Extreme  
from Donostia  
Göttingen, June 22<sup>nd</sup>, 2018



MEMBERS OF THE THESIS COMMITTEE:

Prof. Dr. Jochen S. Hub  
Theoretical Physics

Prof. Dr. Markus Zweckstetter  
Max Planck Institute for Biophysical Chemistry

Dr. Sebastian Kruss  
Institute for Physical Chemistry  
Faculty of Chemistry

ADDITIONAL MEMBERS OF THE THESIS COMMITTEE:

Prof. Dr. Jörg Enderlein  
III. Institute of Physics

Prof. Dr. Claudia Steinem  
Institute of Organic and Biomolecular Chemistry  
Faculty of Chemistry

Prof. Dr. Bert de Groot  
Max Planck Institute for Biophysical Chemistry

**Date of oral examination:** August 7<sup>th</sup>, 2018.



---

DECLARATION

---

I hereby declare that this doctoral thesis entitled "Computational Studies of Small Molecule Permeation across Membrane Channels" has been written independently with no other sources and aids than those quoted.

*Göttingen, June 22<sup>nd</sup>, 2018*

---

Igor Ariz Extreme



*It is the artist who is truthful and it is photography which lies, for in reality  
time does not stop*

— Auguste Rodin (Paris, 1840 - Meudon, 1917)

---

## ACKNOWLEDGEMENTS

---

I am very grateful to my supervisor, Jochen, who has gently guided me throughout these years. Thanks for offering me your time, knowledge and experience, always with a smile. I also want to thank the rest of the Computational Molecular Biophysics Group members, without which the ride would have been very lonely.

This thesis would not have been possible without the economical support of the GGNB, the department of Structural Biology, and the SFB 803 project.

I want to specially thank my partner, Júlia, for living this little adventure beside me.

Finally, I also want to thank my mother, Mila, and my sister, Oihana, as well as the rest of my family and friends, who have always supported me unconditionally. As part of the first generation of my family to study at the University, I want to thank from the bottom of my heart all those women and men, *whose names are unknown*, who have fought through the ages for my right, the right of a member of the working class, to a high quality public education.

Because you were, we are.





---

## ABSTRACT

---

Membrane channels are an essential part of any life form. They conduct the selective flux across the cell membrane of many important molecules that would otherwise not permeate. Experimental studies on membrane channels have led to the structural and functional characterization of many of them, yet many underlying physico-chemical mechanisms are somewhat out of reach. The aim of this thesis is to gain quantitative understanding on the structural and functional properties of these proteins by means of computational methods, such as Molecular Dynamics (MD) and free energy calculations. One of the most common approaches to study the selectivity and permeation mechanisms of a channel is the calculation of the Potential of Mean Force (PMF) for solute permeation across the pore. Usually, PMFs are calculated via MD simulations, which requires a significant amount of computational power. Hence, we compared the capability of MD with that of 3-Dimensional Reference Interaction-Site Model (3D-RISM), allegedly as accurate as MD but much more computationally efficient, to compute PMFs of solute permeation across Urea Transporter B (UT-B) and Aquaporin 1 (AQP1). We found a remarkable agreement between the PMFs for water permeation calculated from both techniques. However, for the rest of tested solutes, namely ammonia, urea, molecular oxygen, and methanol, we found critical discrepancies between 3D-RISM and with MD, which were found to be independent of the closure relation, the choice of the reaction coordinate, or the fluctuations of the protein. This suggests that, whilst 3D-RISM may provide reasonable approximations on PMFs for the permeation of water, it is not appropriate to study the permeation of uncharged non-water solutes.

We further investigated, via a combination of MD simulations and free energy calculations, the structure and function of the fluoride-specific channel Fluc-Bpe. The free energy calculations allowed us to ascertain the specific nature of five isolated electron densities found in the crystal structure of Fluc, four of which were provisionally assigned to fluoride, and the remaining one to sodium. We conducted two different kinds of binding free energy calculations: i) *relative* binding free energy differences  $\Delta\Delta G_{\text{bind}}$ , and ii) *absolute* binding free energy  $\Delta G_{\text{bind}}$ . Notably, the calculation of  $\Delta\Delta G_{\text{bind}}$  allowed us to determine, between two putative molecular species, namely water and fluoride, which species was more likely to bind at a certain binding site. The resulting free energies were partly dependent on fluoride-phenylalanine interactions, which we found to be underestimated by  $\sim 30 \text{ kJ mol}^{-1}$  in current additive force-fields. Thus, the discrimination of one species over the other was only possible because the  $\Delta\Delta G_{\text{bind}}$

values largely deviated from zero. In turn, the calculation of  $\Delta G_{\text{bind}}$  allowed us to confirm whether a certain species would bind *per se* to Fluc-Bpe. Besides, short, free MD simulations proved to be key to assess the structural stability of the channel in different conditions, which, together with the free energy calculations, indicated that the four densities assigned to fluoride rather corresponded to ordered water molecules, and that the last electron density corresponded to a structural sodium.

We finally evaluated, using MD simulations, the response of Fluc-Bpe to the presence of fluoride ions restrained at the permeating pore. The results suggested that the channel would undergo an opening transition, after which water molecules enter the pore to solvate the ions. Then, we calculated the PMFs for the permeation of water, fluoride and chloride using Umbrella Sampling (US) simulations. The profiles of solute permeation across the open structure indicated that water, fluoride, chloride would efficiently permeate the channel, being in stark contrast with the experimental evidence, which demonstrates that Fluc channels permeate fluoride by a  $\sim 100$ -fold ratio over chloride. We suspect that our results might be affected by the inaccurate modelling of ion-protein contacts highlighted before. The proper modelling of ion-protein interactions is extremely important for the establishment of salt-bridges, the structural stability of proteins, or the permeation of ions. Therefore, we conclude that our results regarding the permeation mechanism in Fluc-Bpe mainly reflect the imperfections of current additive force-fields, and that the usage of polarizable force-fields and development of accurate ion-protein interactions may certainly aid future research.

---

## CONTENTS

---

<b>I</b>	<b>INTRODUCTION</b>	<b>1</b>
1	MEMBRANE PROTEINS	3
1.1	Membrane proteins transfer molecules across membranes	3
1.2	Proteins function in motion . . . . .	4
1.3	Protein dynamics can be studied computationally . . .	6
1.3.1	Different methods allow different levels of resolution . . . . .	6
1.3.2	Membrane protein simulations have their own limitations . . . . .	7
2	UREA TRANSPORTERS	9
2.1	UT-B mediates the transport of urea, water and ammonia	9
2.2	The trimeric structure of UT-B . . . . .	9
2.3	S <sub>m</sub> h-bond pattern is crucial in the urea permeation mechanism . . . . .	10
3	FLUC CHANNELS	13
3.1	Fluc channels: classification and physiological role . . .	13
3.1.1	Fluoride is an ubiquitous toxic ion . . . . .	13
3.1.2	The expression of Fluc channels is controlled by a fluoride-responsive riboswitch . . . . .	14
3.1.3	Fluc channels are fluoride specific ion channels	14
3.2	Fluc channels: structural considerations . . . . .	15
3.2.1	Fluc family is topologically diverse . . . . .	15
3.2.2	Bacterial Fluc channels arrange in an antiparallel homodimeric structure . . . . .	15
3.2.3	The double-barreled structure points to a multi-ion single-file permeation mechanism . . . . .	17
3.2.4	Non-functional mutants offer clues about the permeation mechanism . . . . .	19
4	MOTIVATION	23
<b>II</b>	<b>METHODS</b>	<b>25</b>
5	COMPUTATIONAL THEORY AND METHODS	27
5.1	Molecular Dynamics . . . . .	27
5.1.1	Born-Oppenheimer approximation . . . . .	27
5.1.2	Classical MD approximation . . . . .	28
5.1.3	Empirical Force-Fields . . . . .	29
5.1.4	MD algorithms . . . . .	31
5.2	Potential of Mean Force . . . . .	33
5.2.1	Umbrella Sampling . . . . .	34
5.3	Thermodynamic Integration . . . . .	34
5.4	3-Dimensional Reference Interaction-Site Model . . . .	35

5.4.1	3D-RISM solution can be translated into a PMF	37
6	COMPUTATIONAL DETAILS	39
6.1	System set-up for US simulations of UT-B . . . . .	39
6.1.1	US simulation details . . . . .	40
6.1.2	PMF construction after US . . . . .	41
6.2	DRISM and 3D-RISM calculation details . . . . .	42
6.2.1	PMF construction after 3D-RISM . . . . .	43
6.3	IC <sub>50</sub> estimate . . . . .	44
6.4	Water occupancy inside the channel from a PMF . . . . .	44
6.5	System set-up for MD simulations with Fluc-Bpe . . . . .	45
6.5.1	Amber ff99SB*-ILDN force-field . . . . .	45
6.5.2	US simulation details . . . . .	47
6.5.3	PMF construction after US . . . . .	47
6.5.4	CHARMM36 force-field . . . . .	47
6.5.5	Polarizable DRUDE force-field . . . . .	48
6.6	Free-energy calculations of absolute and relative binding in Fluc-Bpe . . . . .	49
6.6.1	Relative binding free-energy computational details . . . . .	51
6.6.2	Absolute binding free energy computational details . . . . .	51
6.7	Maximum-likelihood estimates . . . . .	53
6.8	Quantum chemical calculations . . . . .	53
	<b>III RESULTS</b>	55
7	A COMPARISON BETWEEN MD AND 3D-RISM	57
7.1	PMF Calculations from Umbrella Sampling . . . . .	57
7.1.1	UT-B . . . . .	57
7.1.2	AQP1 . . . . .	59
7.2	PMF Calculations from 3D-RISM . . . . .	61
7.2.1	3D-RISM PMFs for water permeation are consistent with US PMFs . . . . .	61
7.2.2	The solute size dominates the 3D-RISM PMFs . . . . .	62
7.2.3	Structural fluctuations strongly affect PMFs from 3D-RISM . . . . .	63
7.2.4	Alternative closure relations do not improve PMF calculations from 3D-RISM . . . . .	63
7.2.5	Different solvent sites in 3D-RISM yield similar PMFs . . . . .	67
7.2.6	More discrete grid spacing does not change the PMFs from 3D-RISM . . . . .	68
7.2.7	Urea COM <i>vs.</i> C-atom and the LJ parameters on H-atoms do not affect PMFs from US . . . . .	68
8	ASSIGNING THE CRYSTALLOGRAPHIC ELECTRON DENSITIES IN FLUC	71

8.1	Water is more stable than $F^-$ inside Fluc-Bpe in free MD simulations . . . . .	71
8.2	Difference in $\Delta\Delta G_{\text{bind}}$ between $F^-$ and water to Fluc favors water over $F^-$ . . . . .	73
8.3	Absolute binding free energy $\Delta G_{\text{bind}}$ of water to Fluc . . . . .	76
8.4	Rationalizing the disappearance of electron densities in the F82I/F85I mutant. . . . .	76
8.5	TM3 site is stabilized by $Na^+$ . . . . .	77
8.6	The role of the force-field in modeling ion-phenylalanine interactions . . . . .	78
9	THE PERMEATION MECHANISM IN FLUC-BPE . . . . .	83
9.1	Water enters Fluc-Bpe when $F^-$ ions are restrained in the channel . . . . .	83
9.2	PMFs of solute permeation across an open Fluc-Bpe . . . . .	86
IV	DISCUSSION AND CONCLUSIONS . . . . .	89
10	DISCUSSION . . . . .	91
10.1	Is 3D-RISM more suitable than MD to calculate PMFs of solute permeation? . . . . .	91
10.2	The nature of the crystallographic electron densities in Fluc-Bpe . . . . .	94
10.3	The permeation mechanism in Fluc-Bpe . . . . .	96
11	CONCLUSION . . . . .	101
V	APPENDIX . . . . .	103
A	ENTROPY IN A FLAT-BOTTOMED CYLINDRICAL POTENTIAL . . . . .	105
B	TIME TRACES OF $F^-$ AND WATER POSITIONS INSIDE FLUC-BPE . . . . .	107
C	TIME TRACES OF THE AVERAGE NUMBER OF WATER MOLECULES INSIDE FLUC-BPE . . . . .	111
D	ABOUT THE LATERAL DISTRIBUTION OF CHOLESTEROL IN BINARY LIPID MIXTURES . . . . .	115
D.1	Introduction . . . . .	115
D.2	Materials and Methods . . . . .	116
D.2.1	System Set-up for MD simulations . . . . .	116
D.2.2	Pulling simulation details before US . . . . .	116
D.2.3	US simulation and PMF calculation details . . . . .	117
D.2.4	Loss of lateral entropy calculations . . . . .	118
D.2.5	Calculation of $\Delta W_{\mu}(R)$ contributions . . . . .	118
D.3	Results and Discussion . . . . .	119
D.4	Conclusion . . . . .	121
	BIBLIOGRAPHY . . . . .	123

---

LIST OF FIGURES

---

Figure 2.1	Crystal structure of bovine UT-B . . . . .	10
Figure 2.2	Urea permeation across UT-B . . . . .	11
Figure 3.1	Topological diversity in Fluc family channels .	16
Figure 3.2	Crystal structure of Fluc-Bpe . . . . .	18
Figure 3.3	Suggested permeation mechanism in Fluc chan- nel . . . . .	19
Figure 3.4	Detail on electron density disappearance in crys- tal structures of Fluc-Ec2 F80I and F83I mutants	20
Figure 3.5	Residues of Fluc-Ec2 polar-track . . . . .	21
Figure 6.1	US simulation system box with UT-B and 3D- RISM calculation grid . . . . .	40
Figure 6.2	Simulation system box with Fluc-Bpe . . . . .	46
Figure 6.3	Alchemical transformation thermodynamic cycle	50
Figure 6.4	Absolute binding free energy thermodynamic cycle . . . . .	52
Figure 7.1	PMFs from US and 3D-RISM for UT-B . . . . .	58
Figure 7.2	PMFs from US and 3D-RISM for AQP1 . . . . .	60
Figure 7.3	PMFs for water permeation across UT-B and AQP1 . . . . .	61
Figure 7.4	Influence of structural fluctuations on PMFs from 3D-RISM . . . . .	64
Figure 7.5	PMFs from 3D-RISM with KH and PSE-n clo- sures . . . . .	65
Figure 7.6	PMFs from 3D-RISM with KH, PSE-n, and HNC	66
Figure 7.7	PMFs from 3D-RISM using different solvent sites	67
Figure 7.8	PMFs from 3D-RISM using different grid spac- ings . . . . .	68
Figure 7.9	PMFs from US using the C-atom of urea and LJ parameters for H-atoms . . . . .	69
Figure 8.1	F <sup>-</sup> and water positions at the F82 and F85 sites in Fluc . . . . .	72
Figure 8.2	Summary of free-energy calculations . . . . .	74
Figure 8.3	Water and F <sup>-</sup> at F82 and F85 sites, taken from snapshots of the relative binding free energy simulations . . . . .	75
Figure 8.4	On the disappearance of electron densities at the F82I and F85I in the mutant crystals . . . . .	78
Figure 8.5	G77 and T80 residue positions in MD simulations	79
Figure 8.6	Relaxed PESs of ion-benzene interactions . . . . .	80
Figure 9.1	The structure of Fluc-Bpe when the pores are filled with water molecules . . . . .	84

Figure 9.2	Number of water molecules inside Fluc-Bpe . . .	85
Figure 9.3	PMFs of solute permeation across Fluc-Bpe . . .	86
Figure 10.1	Urea binding site in UT-B $S_0$ region . . . . .	93
Figure B.1	$F^-$ positions in Fluc-Bpe blocked by L2 monobodies . . . . .	107
Figure B.2	Water molecule positions in Fluc-Bpe blocked by L2 monobodies . . . . .	107
Figure B.3	$F^-$ positions in Fluc-Bpe free of monobodies . . .	108
Figure B.4	Water molecule positions in Fluc-Bpe free of monobodies . . . . .	108
Figure B.5	$F^-$ positions in Fluc-Bpe free of monobodies simulated with the CHARMM36 . . . . .	109
Figure B.6	$F^-$ positions in Fluc-Bpe free of monobodies simulated with the CHARMM-Drude . . . . .	109
Figure C.1	Average number of water molecules inside Fluc-Bpe when $F^-$ is restrained at F82 and F85 (Amber99SB) <sup>111</sup>	
Figure C.2	Average number of water molecules inside Fluc-Bpe when $F^-$ is restrained at F82 and F85 (CHARMM36) <sup>112</sup>	
Figure C.3	Average number of water molecules inside Fluc-Bpe when $F^-$ is restrained at F82 and F85 (CHARMM36) <sup>112</sup>	
Figure C.4	Average number of water molecules inside Fluc-Bpe when water is bound to F82 and F85 . . .	113
Figure D.1	CG representation of water and lipids in Martini FF . . . . .	116
Figure D.2	CG system in pulling simulations . . . . .	117
Figure D.3	PMFs $W_D(R)$ for cholesterol-depleted domain formation . . . . .	120

---

LIST OF TABLES

---

Table 6.1	Closures and tolerances in DRISM calculations	54
Table 6.2	Closures and tolerances in 3D-DRISM calculations . . . . .	54
Table 7.1	Average water occupancy in UT-B and AQP1 .	62
Table 8.1	Maximum-likelihood estimates $\tau_{ml}$ (nanoseconds) for the lifetime of $F^-$ and water molecules at Fluc-Bpe F82 and F85 sites . . . . .	73
Table 8.2	Absolute binding free energy $\Delta G_{bind}$ ( $\text{kJ mol}^{-1}$ ) of water at F82, F85, F82I, and F85I sites . . . .	76
Table 9.1	Number of water molecules in Fluc-Bpe pores	85



---

## ACRONYMS

---

IDP	Intrinsically Disordered Protein
NMR	Nuclear Magnetic Resonance
QM	Quantum Mechanics
MM	Molecular Mechanics
MD	Molecular Dynamics
CG	Coarse-Grained
DFT	Density Functional Theory
UT	Urea Transporter
UT-A	Urea Transporter A
UT-B	Urea Transporter B
UT-C	Urea Transporter C
WT	wild-type
AQP <sub>1</sub>	Aquaporin 1
PDB	Protein Data Bank
PMF	Potential of Mean Force
US	Umbrella Sampling
<sub>3</sub> D-RISM	3-Dimensional Reference Interaction-Site Model
TI	Thermodynamic Integration
LJ	Lennard-Jones
PME	Particle-Mesh Ewald
SCF	Self Consistent Field
WHAM	Weighted Histogram Analysis Method
OZ	Ornstein-Zernike
RISM	Reference Interaction-Site Model
DRISM	Dielectrically consistent Reference Interaction-Site Model
PY	Percus-Yevick

HNC	HyperNetted-Chain
MSA	Mean-Spherical Approximation
KH	Kovalenko-Hirata
PSE- <i>n</i>	Partial Series Expansion of order- <i>n</i>
COM	center-of-mass
POPC	1-palmitoyl-2-oleoyl- <i>sn</i> -glycero-3-phosphocholine
DPPC	1,2-dipalmitoyl- <i>sn</i> -phosphatidilcholine
DOPC	1,2-dioleoyl- <i>sn</i> -glycero-3-phosphocholine
PES	Potential Energy Surface
MP <sub>2</sub>	Møller-Plesset 2 <sup>nd</sup> -order
ar/R	aromatic/arginine
RMSF	Root Mean Squared Fluctuations
RMSD	Root Mean Squared Deviation
RDF	Radial Distribution Function
CRAC	Ca <sup>2+</sup> release-activated Ca <sup>2+</sup>

Part I

INTRODUCTION



---

## MEMBRANE PROTEINS

---



THE lipidic barriers commonly known as 'cell membranes' not only constitute the separation between two spatial realities, i. e. the inner- and outer-cell spaces, but also delimit the permanent exchange of molecules between those. Cell membranes are characterized by their selective permeability, which keeps the dynamic electrochemical balance between the inner and outer species. Whilst hydrophobic molecules passively move across membranes by diffusion, polar molecules need, in most cases, the mediation of protein channels and transporters. The understanding, in atomic detail, of the dynamics of such proteins can not only improve our basic understanding of their function but also lead to the design of molecular filters with implications in the health and chemical industry. To unveil the permeation mechanisms of membrane proteins, usually a combination of experimental and computational methods is required. In this thesis we used the latter to study the structural and thermodynamic properties of two different families protein channels, namely, the Urea Transporters and the Fluc channels.

### 1.1 MEMBRANE PROTEINS TRANSFER MOLECULES ACROSS MEMBRANES

Cell membranes play a crucial role in all forms of life. They must meet three main requirements: first, to constitute a separation between the cellular volume and the outside, thereby keeping a distinct electrochemical composition between the cell interior and the outer environment. Second, they have to allow the entering of certain substances such as nutrients and ions into the cell, and the expulsion of waste compounds outwards. Third, cell membranes need to be physically flexible enough to allow cells to grow and change shape. Cell membranes leverage the physical properties of the molecules that constitute them to be able to adjust to these multiple needs [1]. To allow the efficient entering and exiting of certain compounds while the rest remain appropriately inside or outside the cell is the characteristic known as 'selective permeability'. This accurate and essential mechanism of exchange is usually driven by proteins, which account for a large fraction of mass of the membrane [2, 3].

*An insightful summary of protein to phospholipid mass ratios is shown in Mitra et al. (2004) [2].*

For each of the different compounds that cross the membrane the permeability is defined as

$$\text{flux} = P(c_{\text{in}} - c_{\text{out}}) \quad (1.1)$$

where the flux is the number of molecules crossing a unit area per unit time, the permeability coefficient  $P$  has units of length/time, and the  $c_{\text{in}}$  and  $c_{\text{out}}$  are the concentrations of the compounds in and out of the cell measured in number of molecules per unit volume [1]. Some membrane proteins selectively increase the permeability for certain molecules allowing them to move down its concentration gradient. Such proteins are known as 'channels', and some of them adopt a so called 'open-state' conformation, allowing molecules in and out constantly. Other channels are gated, and cycle between open- and close-states, opening only when a particular signal is received like, for example, the binding of a ligand, a change in trans-membrane potential, or a mechanical signal. Another kind of membrane proteins needs energy to actively transport molecules against their concentration gradient. These are known as 'transporters' or 'pumps' and they accumulate nutrients such as sugar and amino acids from the surroundings, playing also a critical part in creating ionic concentration gradients and keeping them out of equilibrium [4].

The boundary between channels and pumps is somewhat blurry [5, 6] and, in addition to the direction of the transport—with channels following the concentration gradient and pumps going against the concentration gradient—, they can also be distinguished by the unitary flux rate observed in experiment. Accordingly, channels are characterized by higher flux rates (with unitary flux rates greater than  $10^6 \text{ s}^{-1}$ ) than transporters (with unitary flux rates between 1–1000  $\text{s}^{-1}$ ). The flux rates translate into different free energy barriers which correspond to different limiting steps. These are related, in turn, to two types of transport, namely low energy barriers for diffusion (open channels), and high energy barriers for conformational changes (close-to open-state gating events).

As we will see, to fully understand the underlying mechanisms of a given transport process, an ensemble of atomic-resolution structures is needed. This thesis describes computational studies of Urea Transporters and the Fluc family channels. The formers are known for mediating the passage of several molecules such as urea, water, or ammonia across the cell membrane in a wide range of species. Fluc are fluoride specific channels involved in the resistance to fluoride toxicity (see Chapter 2 and Chapter 3).

## 1.2 PROTEINS FUNCTION IN MOTION

The conformational dynamics of protein channels and transporters are needed to undergo a transition from a close- to an open-state and

*The Transporter Classification Database <http://www.tcdb.org/> provides a thorough listing and a lot of useful information about channels and pumps [7].*

to allow, thereby, substrate pass from one side of the cell to the other. Yet membrane channels and pumps are no exception, since all proteins develop their functions in motion. The static view of proteins where a steady structure was the main responsible of its function has been long overcome [8]. Most proteins exhibit a stable 3-dimensional fold, but to perform their functions they move, thereby exploiting their structural plasticity to sample different conformers. The ensemble of thermally accessible conformational states that proteins sample can be expressed as a multi-dimensional energy landscape that establishes their relative probabilities and the free energy barriers between them [9]. In addition, there is a vast amount of proteins that do not show a stable folded structure. They are the so-called IDPs, and their conformational free-energy surface is much flatter than that of folded proteins, allowing IDPs to seize on functional mechanisms inaccessible to proteins with a 3D fold [10, 11]. This theoretical frame, in which the relation between protein conformers and dynamics is seen as a multi-dimensional energy landscape, has helped to shed light onto many important biological processes such as molecular recognition, signaling, catalysis, gene regulation, protein aggregation, protein folding, ligand binding, or functional selection in evolution [12–16].

Critical processes such as protein-protein interactions, catalysis, or signal transduction occur in the microsecond time-scale and involve excited conformations separated by energy barriers of several  $k_B T$  (the product of the Boltzmann constant and the absolute temperature). Within this ‘slow’ time-scale, fast small-amplitude picosecond-to-nanosecond dynamics happen at physiological temperature. These fast thermal fluctuations, that range from bond vector vibrations to side-chain rotations, can be sampled from a large ensemble of structurally similar ground-state conformations separated by energy barriers of less than  $1 k_B T$  [17]. Such ground-state conformations that occupy the bottom of the energy landscape have been the subject of a wide variety of experimental studies by Nuclear Magnetic Resonance (NMR), hydrogen exchange, or time-resolved X-ray structure analysis [18–22]. On the other hand, the excited-state conformations that are separated by several  $k_B T$  barrier heights have been out of reach of the techniques mentioned above in the past. Fortunately, the progress made in solution NMR spectroscopy has made it easier to study such excited states [23]. All these experimental techniques and others such as cryo-electron microscopy, small- and wide-angle X-ray scattering, and fluorescence single-molecule methods can be used and combined to study a wide range of protein dynamics in different time-scales and resolutions. However, whereas an increase in resolution usually entails a decrease in dynamical information [17], computational methods reveal themselves as an essential sidekick to experimental techniques, offering atomic-detail resolution as well as a full dynamic picture.

*Jensen et al. (2014) wrote a thoroughful piece of work about how to study IDPs within this conceptual framework [10].*

### 1.3 PROTEIN DYNAMICS CAN BE STUDIED COMPUTATIONALLY

The main caveat of experimental techniques is that they cannot resolve the motion of individual atoms moving within a protein. Fortunately, computational biophysical methods may be used to simulate protein dynamics in atomic detail. The simulated dynamics will then determine the phase space trajectories of the protein molecule at a given temperature, thus providing information about its thermodynamic properties. Protein motions along their free-energy landscape can be computed as a function of time according to different physical models and mathematical approximations, in which three main resolution levels can be distinguished, from the most accurate and expensive from a computationally, to the most simplistic and most computationally efficient: namely, i) Quantum Mechanical (QM) description, ii) classical Molecular Mechanics (MM) representation, also known as MD, and iii) Coarse-Grained (CG) representation.

#### 1.3.1 *Different methods allow different levels of resolution*

QM methods calculate the movements along time of nuclei and electrons by solving the time-dependent Schrödinger equation (Equation 1.2) with different approximations:

$$H\psi = i\hbar \frac{\partial \psi}{\partial t} \quad (1.2)$$

where  $H$  is the Hamiltonian operator,  $\hbar$  is the Planck constant  $h$  divided by  $2\pi$ , and  $\psi$  is the molecular wave function. QM methods provide the highest level of detail but they are, at the same time, the most expensive in terms of computational resources, which i) has limited its application to small systems of isolated molecules in gas phase or molecules embedded in continuum solvents, and ii) has fueled the development of alternate and simpler methods like the ones based on the Hartree-Fock method or the Density Functional Theory (DFT) [24].

The next level of detail comprises MD methods. In these, molecules are represented in atomic-detail, and the atoms are, in turn, represented as point-charges and Lennard-Jones (LJ)-spheres. A classical force-field (Equation 1.3), that has been parametrized based on high quality QM calculations and experiments, calculates the forces on every atom from its first derivative with respect to the atom positions. Newton's equations of motion are then integrated to calculate the change of atom positions in time [25]. The basic functional form of a classical force-field  $V(\mathbf{R})$  is the following:

$$V(\mathbf{R}) = \underbrace{V_{\text{bonds}} + V_{\text{angles}} + V_{\text{dihedrals}} + V_{\text{imp. dih.}}}_{\text{bonded}} + \underbrace{V_{\text{LJ}} + V_{\text{coulomb}}}_{\text{non-bonded}} \quad (1.3)$$



where the potential with respect the atom positions  $V(\mathbf{R})$  is the sum of bonded and non-bonded interactions (see Chapter 5 for a more detailed description). Since the first simulation by McCammon *et al.* (1977) [26], MD simulations have been used to simulate a wide range of protein systems and macromolecules in explicit solvent. Experimental techniques are still needed to parametrize force-fields and to obtain the starting structures for simulations, but MD simulations have proven to be able to solve questions that have been impossible to answer with experiments, like the pathway between the open and close conformations of GroEL1, or the dynamic functional features in acetylcholinesteras [27, 28]. Furthermore, the limitations that are yet inherent to the method have been overcome gradually. On one hand, force-fields have been subjected to a large improvement in parametrization and have evolved over the last years, allowing, for example, to simulate more accurate folding dynamics, IDP dynamics, or to include polarization effects [29–31]. On the other hand, the development of algorithms and computer hardware has increased the sampling capabilities of MD simulations, breaking the milisecond barrier [33–35].

In the lowest level of resolution covered in this introduction, CG methods, the full atomic detail is lost, thereby reducing the computational cost. Although the force-field functional forms can resemble the MD ones, they use non-physical statistical potentials that have to be carefully assessed to reproduce accurate structural properties of proteins. One of the most used CG methods is the MARTINI force-field, in which four heavy atoms are represented with a single bead [36]. Many researchers have contributed to this method, improving its accuracy in simulations with proteins, carbohydrates, and nucleic acids [37–39].

### 1.3.2 Membrane protein simulations have their own limitations

In the particular case of membrane proteins, there have not been as many computational studies as with solvated proteins, because it has traditionally been difficult i) to obtain high quality membrane protein structures [41], and ii) to find accurate parameters to reproduce the thermodynamical properties of the lipids that form the membrane bilayers [42, 43]. Fortunately, nowadays we enjoy a fair amount of computational tools aimed at studying membrane proteins, with which multi-lipid membranes can be generated, we can embed proteins in these membranes or develop computational electrophysiology studies [44–46]. Such advances have so far helped to elucidate channel activation and ion permeation mechanisms, substrate translocation mechanisms, the role of lipid interactions with membrane proteins, to carry out the above-mentioned computational electrophysiology studies, or even to design *de novo* channels [47–52].

*Yu et al. (2016) carried out a 20 ns simulation of the crowded environment of the cell with a 100 million atom system [32].*

*The Membrane Proteins of Known 3D Structure database <http://blanco.biomol.uci.edu/mpstruc/> is an invaluable resource for finding high quality structures of membrane proteins [40].*



---

## UREA TRANSPORTERS

---



FAMILY of membrane proteins known as Urea Transporters (UTs) mediate the flux of small, polar molecules across the cell membrane. Originally, UTs were predicted to be transporters, but since single-channel flux rates of  $10^4$ - $10^6$  urea molecules per second were measured, they started to be considered as channel-like proteins [53, 54]. UTs can be divided further into three subfamilies, namely UT-A, UT-B, and UT-C. The first two subfamilies are widespread in the biota, having been detected in multiple animal and bacterial species, whereas UT-C has been characterized in only two species of fish [55]. In this chapter we summarize the functional and structural characteristics of UT-B, a channel extensively studied both experimentally and computationally.

### 2.1 UT-B MEDIATES THE TRANSPORT OF UREA, WATER AND AMMONIA

UT-B is encoded by the gene *Slc14a1*, which is mainly expressed in erythrocytes, being common also in kidney, brain, ear, testis, intestine and bladder [56–60]. The main function of UT-B is to mediate the transport of urea, water and ammonia following their concentration gradients. Experiments with UT-B knockout mice showed that the absence of UT-B in kidney results in ‘a urea selective urinary concentrating defect’ compared with wild-type (WT) mice [61]. Also, Yang & Verkman (2002) showed that i) the permeability of water in erythrocytes of AQP<sub>1</sub>/UT-B knockout mice was significantly lower as compared with that of AQP<sub>1</sub> knockout mice, and ii) that the single channel water permeability of UT-B was comparable to that of AQP<sub>1</sub> [62]. In addition, a combination of experimental techniques and MD simulations demonstrated that NH<sub>3</sub> gas can permeate UT-B [63].

### 2.2 THE TRIMERIC STRUCTURE OF UT-B

The crystal structure of the bovine UT-B was solved at a resolution of 2.36 Å (PDB IDs: 4EZC, 4EZD). UT-B (Figure 2.1) forms a trimer with a cavity filled with lipid or detergent molecules in the interface of the monomers. Each protomer presents two inverted homologous halves

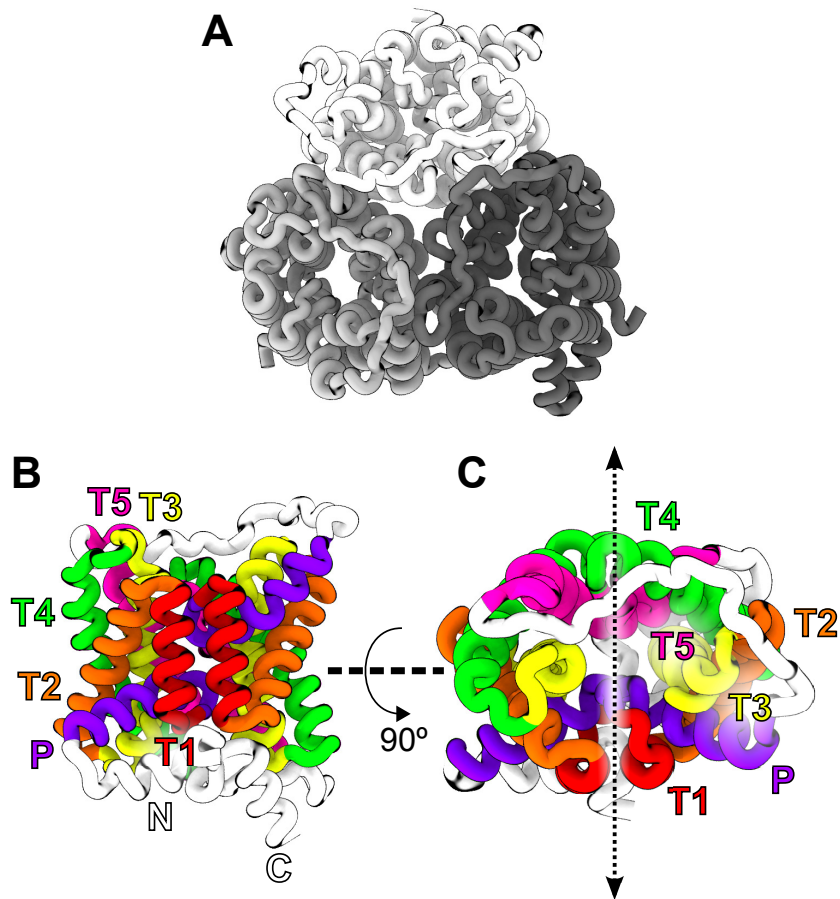


Figure 2.1: **Crystal structure of bovine UT-B.** (A) Top-view of the UT-B trimer. (B) Monomeric structure of UT-B as viewed from a plane parallel to the membrane and (C) from the extracellular side perpendicular to the membrane. The transmembrane helices in B and C are colored in pseudo-symmetry-related pairs.

which are formed, in turn, by five trans-membrane helices (T1a-5a and T1b-5b). The interface of the two halves form the pore, an amphipatic membrane-spanning tunnel in which three regions can be highlighted: the  $S_i$  and  $S_o$  sites, and the  $S_m$  region (Figure 2.2).  $S_i$  and  $S_o$ , lined by carbonyl and side-chain oxygen atoms, connect the *inner*- and *outer*-cell with  $S_m$ . The central region  $S_m$  is a narrow selectivity filter that, save for a pair of threonine residues, is fully hydrophobic [64].

### 2.3 $S_M$ H-BOND PATTERN IS CRUCIAL IN THE UREA PERMEATION MECHANISM

In the crystal, selenourea is bound to  $S_i$  and  $S_o$  sites, for which independent computational studies revealed two free energy minima in the Potential of Mean Force (PMF). These urea binding sites are found at both sides of the  $S_m$  region, in which a free energy maximum is

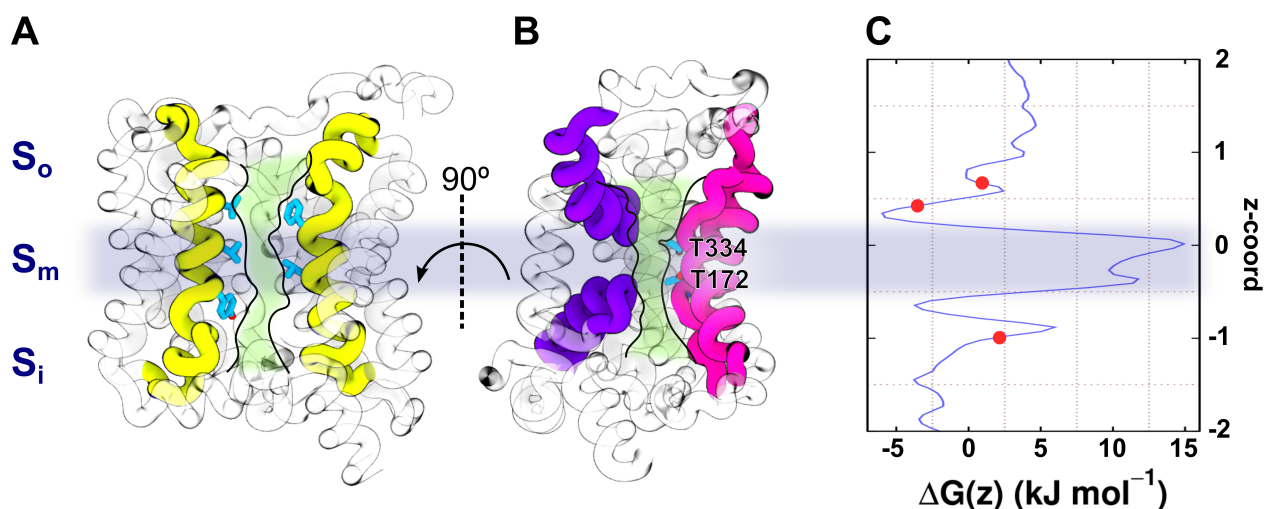


Figure 2.2: **Urea permeation across UT-B.** (A) The urea permeation pathway is shown viewed from a plane parallel to the membrane, with the cross-section of the pore marked in green. The residues T172 and T334 are highlighted by turquoise sticks, and the  $S_o$ ,  $S_m$  regions surround  $S_i$ , highlighted by a blue shade. (C) PMF  $\Delta G(z)$  of urea permeation through the UT-B pore. The red dots indicate the positions where selenourea is bound in the crystal structure. The PMF  $\Delta G(z)$  has been calculated as explained in Section 6.1 [65].

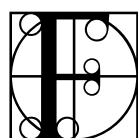
observed [64, 65]. Across this selectivity filter, urea permeates by establishing hydrogen bonds along the oxygen ladder without causing much perturbation to the channel (Figure 2.2). Levin *et al.* (2012) showed experimentally that the T334V mutant, which decreases the number of available hydrogen bonds in  $S_m$ , exhibits reduced urea efflux as compared with the WT, whereas the T172S, T334S, and T172S/T334S mutants, which conserve the hydrogen bonding pattern, had little effect on the rate of urea efflux. Moreover, in equilibrium MD simulations of T172V/T334V mutant, V334 rotated toward the pore, not allowing even water to permeate [64].



---

 FLUC CHANNELS
 

---



FLUC CHANNELS are a family of small fluoride specific membrane proteins. They are expressed in a wide variety of organisms, in which Fluc channels help to maintain the cytoplasmic  $F^-$  concentration below enzyme inhibitory levels providing, thus, protection in fluoride enriched environments. Although the recently discovered X-ray crystal structure and experimental evidence hint at the permeation mechanism, its minutiae remain unknown.

## 3.1 FLUC CHANNELS: CLASSIFICATION AND PHYSIOLOGICAL ROLE

 3.1.1 *Fluoride is an ubiquitous toxic ion*

Fluoride is an inorganic anion that can be easily found in the environment with concentrations ranging from 10 to 100 mM. In addition, due to its bacterial growth inhibition capability, it is widely used in public water supplies and dental products [66, 67]. In acidic environments,  $F^-$  is found in the form of hydrogen fluoride (HF), to which bacterial membranes are highly permeable.  $F^-/HF$  is a toxic, weak acid ( $pK_a = 3.15$ ) that forms aluminofluoride and beryllofluoride complexes, thereby inhibiting enzymes such as nitrogenases and regulatory phosphatases. The electronic properties of  $F^-$  are akin to those of  $OH^-$ , and therefore substitutes hydroxyde or water in some metalloenzymes. Notably,  $F^-/HF$  binds to heme-groups and other metal-containing groups, inhibiting enzymes such as catalase, urease, or enolase, that are essential to many catabolic processes. Besides, HF acts as a transmembrane proton transporter, causing severe physiological disorders in the cell [68–70].

Fluoride toxicity has been a constant threat in evolution. As a consequence, fluoride exporters can be found in many different organisms, clustered into only two families: i)  $CLC^F F^-/H^+$  antiporters, and ii) the Fluc family, also known as CrcB, or FEX in eukaryotes.  $CLC^F$  antiporters are a subclass of the widespread CLC superfamily, a well known group of anion channels and  $Cl^-/H^+$  antiporters that take part in many important biological processes such as skeletal muscle excitability, blood pressure regulation, acid resistance in enteric bac-

teria, and acidification of lysosomes [71–73]. Besides, the Fluc family, which is not related phylogenetically to the CLC family, comprises a set of small membrane proteins that are broadly distributed among unicellular organisms from bacteria, archaea, and eukarya, but can also be found in several multicellular organisms such as plants, and simple marine animals like sponges, anemones or tunicates [74–76].

### 3.1.2 *The expression of Fluc channels is controlled by a fluoride-responsive riboswitch*

Fluc channels —formerly known as the *crcB* family— are expressed upon the binding of  $F^-$  to a fluoride-responsive riboswitch that regulates translation initiation. This riboswitch, a non-coding RNA structure in which the *crcB* conserved motif is present, is located upstream of genes that encode for several proteins, including various ion channels [77, 78]. The addition of NaF causes structural changes in its most conserved nucleotides, suggesting the formation of a fluoride-binding aptamer. Thus, the addition of  $F^-$  to the environment triggers the expression of Fluc channels, leading to bacterial fluoride resistance. For instance, in cell growth experiments, WT *Escherichia coli* cell growth rate is reduced at 30 mM NaF, whereas *crcB* knockout *E. coli* strains are inhibited at much lower  $\mu$ M NaF concentrations [77].

*Although phylogenetically unrelated to Fluc, members of the CLC superfamily are also associated to fluoride-responsive riboswitches [72]. The X-ray crystal structure of such a  $F^-$ -riboswitch has been solved by Ren et al. (2012) [79].*

The number of riboswitches encoded in the genome is associated with the ability to mitigate toxic effects. For instance, whilst most bacteria species encode one or two fluoride-responsive riboswitches, in the genome of *Methylobacterium extorquens* DM4 at least 10 of them are found [77]. *M. extorquens* DM4 is able to grow on dichloromethane thanks to a halogenase enzyme that can catalyze the degradation of several halogenated hydrocarbons [80–82]. This suggests that *M. extorquens* DM4 might also degrade fluorinated hydrocarbons, thereby increasing the cytoplasmic concentration of  $F^-$ , which would require a very sensitive fluoride-responsive mitigation system for survival.

### 3.1.3 *Fluc channels are fluoride specific ion channels*

The selectivity of Fluc channels has been thoroughly studied in electrophysiological experiments. According to  $F^-$  and  $Cl^-$  efflux experiments with liposomes, Fluc channels are highly selective for fluoride over chloride [74]. The electrophysiological experiments in mono- and bi-ionic conditions showed that i) Fluc channels are also selective for  $F^-$  over  $Na^+$ , with  $F^-$  currents that appear as a time- and voltage-independent leaks over a wide range of voltages, and ii) that the  $F^-/Cl^-$  permeability ratio is  $\sim 100$ . Furthermore, single channel current experiments show a persistent ( $> 95\%$  of the time) ‘full-open’ state at membrane potentials in the range of  $-200$  mV to  $+200$  mV in which  $\sim 10^7$   $F^-$  ions/s cross the channel at  $-200$  mV [74]. Hence,



Fluc channels can be defined as  $F^-$ -specific channels through which fluoride moves across by electrodiffusion.

### 3.2 FLUC CHANNELS: STRUCTURAL CONSIDERATIONS

#### 3.2.1 *Fluc family is topologically diverse*

Fluc family members are found in three different topological states (Figure 3.1): i) antiparallel homodimers—which will be thoroughly discussed in the following sections—encoded by single genes, ii) oppositely inserted heterodimers, and iii) two fused Fluc proteins linked by a trans-membrane helix (only in eukaryotes). The fused eukaryotic Flucs always conserve key residues, such as two critically important phenylalanines, which remain only in one of the two possible pores (see Section 3.2.2 and Section 3.2.3). A similar feature is observed in heterodimeric Flucs, although the conserved pore might differ [76, 83]. As can be seen in Figure 3.1, the topological asymmetry developed in fused Flucs and heterodimers has an obvious advantage: the ability to display side-specific features, like the ones already observed in the *Saccharomyces cerevisiae* Fex1p channel [84].

#### 3.2.2 *Bacterial Fluc channels arrange in an antiparallel homodimeric structure*

Before the X-ray crystal structures of two bacterial Fluc channels were solved by Stockbridge *et al.* (2015) [85], the only clues about its arrangement in the membrane were given by size-exclusion chromatography, photobleaching, and 'Poisson-dump' experiments, which revealed Fluc was a  $\sim 32$  kDa homodimer [74]. Besides, double-sided channel-blocking experiments confirmed the antiparallel orientation of Fluc homodimers by which their monobody-binding epitopes are presented to both sides of the membrane simultaneously [86–88]. Such a dual-topology homodimeric and antiparallel architecture has been previously seen in multi-drug transporters such as EmrE [89, 90], but never before in ion channels.

The crystal structure of Fluc channel from *Bordetella pertussis* (hereafter denoted Fluc-Bpe) was obtained by X-ray diffraction with a resolution of 2.17 Å (PDB ID: 5NKQ) [85]. Figure 3.2 shows the antiparallel homodimeric arrangement of Fluc-Bpe, in which each monomer is constituted by four membrane-spanning helices (TM1-TM4). TM3 is, in turn, divided into two halves (TM3a and TM3b) that are connected by a six-residue loop in the middle of the membrane (TM3-break). The channel has the shape of two cones opposed by the vertices, whose edges, with wide-open vestibules, point outwards the membrane. However, no water-filled pore connects the vestibules. Likewise, the vestibules in the crystal structure are not filled by water,

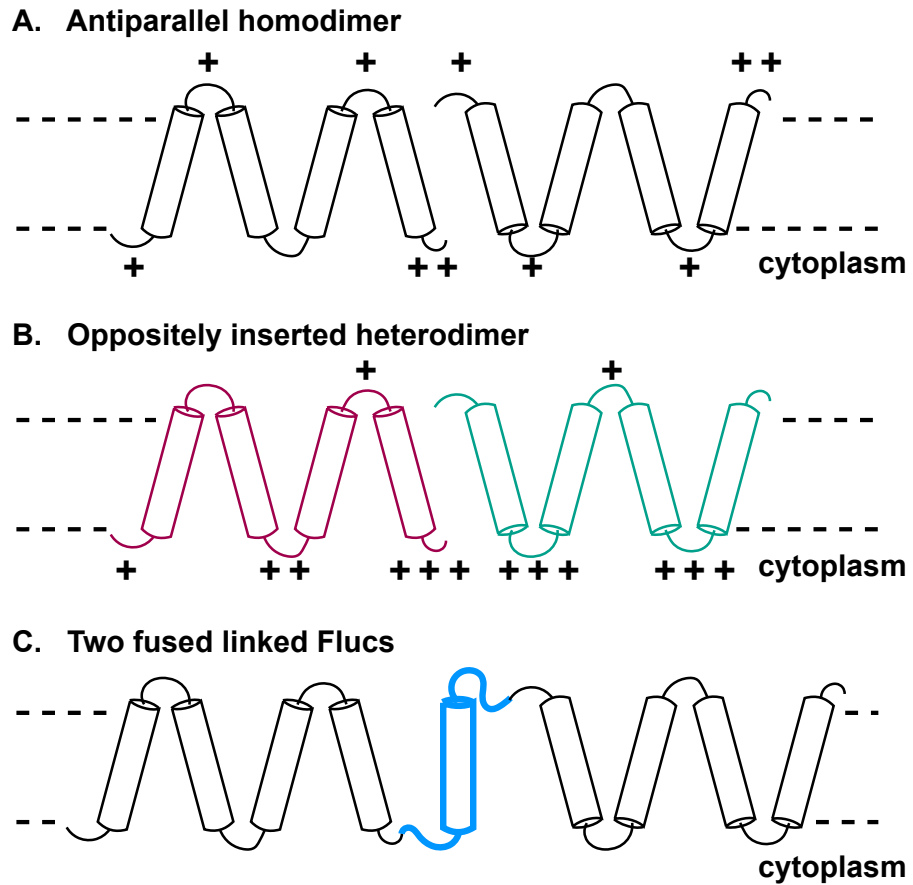


Figure 3.1: **Topological diversity in Fluc family channels.** Schematic view of the different topological states of Fluc family members. The (+) symbols denote arginine and lysine residues. (A) Bacterial antiparallel homodimers are inserted in the membrane in both inward and outward-facing orientations. The Fluc channel from *Bordetella pertussis* is a characteristic example of this topology. (B) Oppositely inserted heterodimers have a fixed orientation. *Lactobacillus acidophilus* is a representative member which shows this topology. (C) Fused Flucs are linked by a transmembrane helix. This kind of topology is more common in eukaryotes such as *Saccharomyces cerevisiae*

but instead, L2 monobodies are bound to the vestibules introducing a long loop of 8 – 10 residues.

The most intriguing part of Fluc-Bpe crystal structure corresponds to five disconnected electron densities: one in the center of the gap connecting the two vestibules, and four at the crannies, which are disconnected from the vestibules, formed between TM2, TM3b, and TM4 (Figure 3.2). The first electron density is coordinated by four backbone carbonyl groups from G77 and T80 residues (two from each monomer) in the TM3-break. Accordingly, it was identified as a structural  $\text{Na}^+$  ion that unavoidably binds during the homodimer assembly process. The remaining set of four disconnected electron densities was provisionally identified as four  $\text{F}^-$  ions which simultaneously occupy the channel (with occupancies of  $\sim 80\%$ ) [85, 91]. They are coordinated by residue N43 from TM2, residues S108 and S112 from TM4, and F82 and F85 from the surroundings of the TM3-break (Figure 3.2). These electron densities are congruent with a halide coordination shell, in which the interaction of  $\text{F}^-$  with the edge of the aromatic rings of F82 and F85 residues might play an important role [92–94]. However, based on X-ray scattering data alone, they cannot be distinguished from ordered water molecules.

In addition to Fluc-Bpe, Stockbridge *et al.* also solved the crystal structure of Fluc channel from *E. coli* (hereafter denoted Fluc-Ec2) by X-ray diffraction with a resolution of 2.6 Å (PDB ID: 5A43). Fluc-Ec2 is blocked by S9 monobodies and shows the same fold as Fluc-Bpe, including the equivalent  $\text{Na}^+$  density, and two out of four similarly coordinated  $\text{F}^-$  densities in similar locations.

### 3.2.3 *The double-barreled structure points to a multi-ion single-file permeation mechanism*

The presence of these four  $\text{F}^-$  ions reveal a channel with a double-barreled structure in which two pores are outlined. Each of the pores is formed by TM2, TM3b, and TM4 of one monomer, plus the phenylalanine from the TM3-break of the opposing monomer. This kind of structure is very uncommon among ion channels, which usually present a single membrane-spanning pore. In the crystal structure, the pores are occupied by two  $\text{F}^-$  ions, suggesting that Fluc-Bpe and Fluc-Ec2 would mediate a multi-ion permeation similar to those of  $\text{K}^+$ -channels that present a single-file of ions aligned in a single pore [95]. The narrow bore of the pore in Fluc channels could be responsible for the selectivity mechanism, which excludes the larger  $\text{Cl}^-$  but not the smaller  $\text{F}^-$ . Besides, the highly conserved residues N43, F82, and F85 may play an important role in  $\text{F}^-$  recognition and permeation, due to short-range quadrupolar interactions with the phenylalanines, or thanks to direct H-bonding to a rotating asparagine that would facilitate the pass of fluoride (Figure 3.3). How-

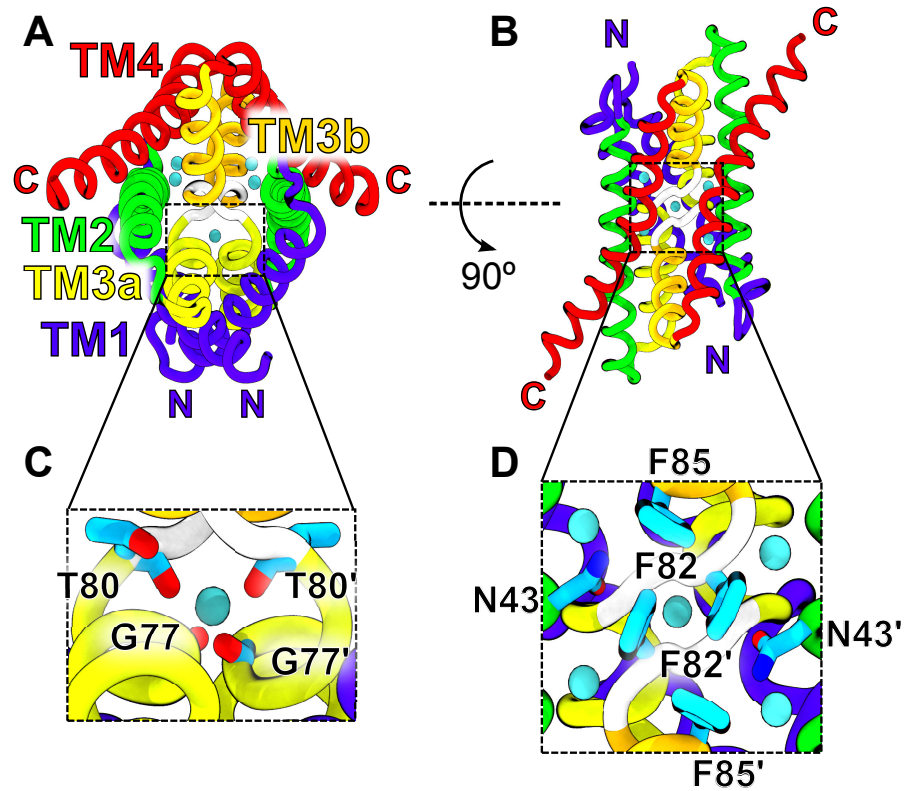


Figure 3.2: **Crystal structure of Fluc-Bpe.** (A) Top-view (perpendicular to the membrane) of the crystal structure of Fluc-Bpe homodimer. Transmembrane helices are coloured as their tags. (B) Side view (parallel to the membrane) of Fluc-Bpe homodimer. (C) Close-up view of the TM3 binding site in which the position of the disconnected electron density can be observed. The coordinating Thr and Gly residues are highlighted. (D) Close-up view of the F82 and F85 binding sites in which their associated disconnected electron densities are also depicted. The coordinating Phe and Asn residues are highlighted.

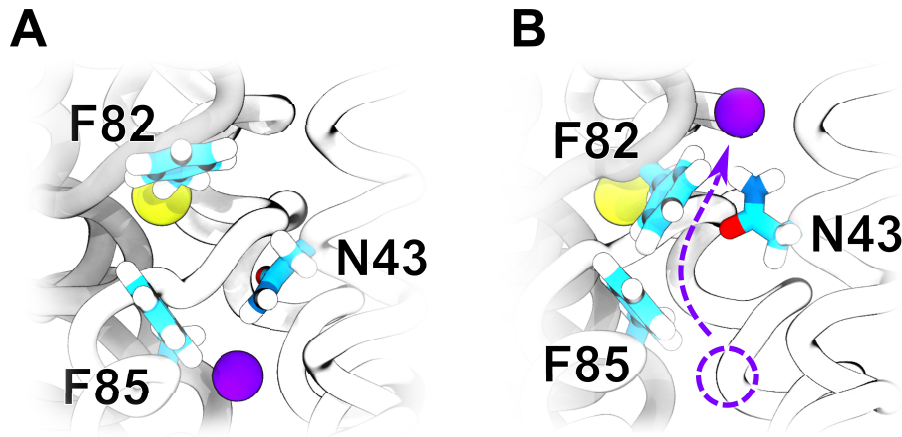


Figure 3.3: **Suggested permeation mechanism in Fluc channel.** (A) Initial position of F<sup>-</sup> coordinated by N43 and F85. (B) N43 rotates and allows F<sup>-</sup> to cross the pore. The arrows indicate the direction of F<sup>-</sup> permeation by electrodiffusion. The structural Na<sup>+</sup> is depicted in yellow.

ever, given the high free energy penalty of desolvation for fluoride ( $\sim 450 \text{ kJ mol}^{-1}$ , *versus*  $\sim 300 \text{ kJ mol}^{-1}$  for chloride [96]), the selective molecular recognition of F<sup>-</sup> in such an anhydrous pore is still an unsolved chemical puzzle [85].

#### 3.2.4 *Non-functional mutants offer clues about the permeation mechanism*

To test the influence of several conserved residues in the functional properties of Fluc channels, mutation experiments were conducted in Fluc-Bpe and Fluc-Ec2. The mutation of the conserved phenylalanines in Fluc-Bpe leads to severely reduced F<sup>-</sup> efflux in the case of F82I and F85I as compared with WT. In addition, the mutant N43D conserves a F<sup>-</sup>-selective efflux at pH 7 that is lost at pH 9 [85]. In turn, the crystal structures of the F80I and F83I mutants of Fluc-Ec2 (PDB ID 5KBN and 5KOM, respectively), show no disconnected electron densities associated to the mutant Ile residues, but retain the other pair of densities coordinated to Phe (Figure 3.4). The Fluc-Ec2 mutant channels, in line with the above-mentioned Fluc-Bpe mutants, are likewise non-functional, revealing a  $> 10^4$ -fold F<sup>-</sup>-efflux inhibition. Furthermore, a constructed concatemer (a Fluc-Ec2 tandem dimer), which in its WT/WT form is functionally indistinguishable from the WT homodimer, allowed to mutate the conserved Phe residues one by one. Otherwise, since the original Fluc-Ec2 channel is a homodimer, the mutation of one residue in the DNA sequence leads to two mutated residues in the translated protein. Therefore, the construction of the concatemer allowed to evaluate single-pore channel-disruptive mutations. Among them, two types of double mutants can be distinguished: i) *trans*, with each mutation in a different pore, and ii) *cis*,

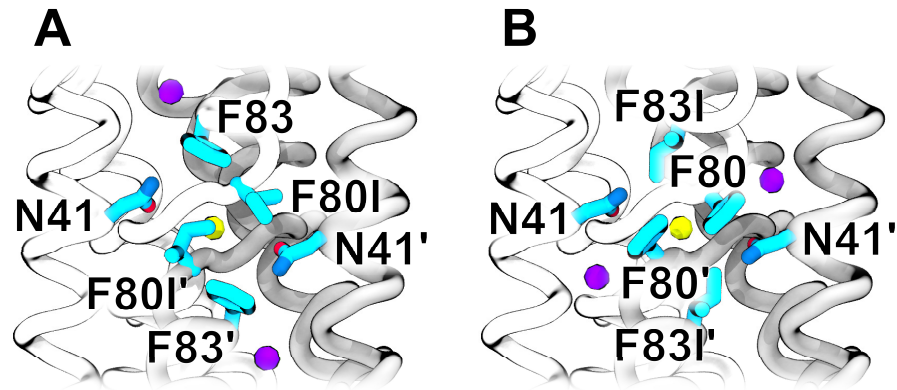


Figure 3.4: **Detail on electron density disappearance in crystal structures of Fluc-Ec2 F80I and F83I mutants.** (A) F80I Fluc mutant (PDB ID: 5KBN). (B) F83I Fluc mutant (PDB ID: 5KOM). The electron densities associated to F80 and F83 sites are coloured in purple, and the structural Na<sup>+</sup> ion is coloured in yellow.

with both mutations sharing a pore. The two *trans* double mutants, F80I/F80I and F83I/F83I, were non-functional, whilst both *cis* double mutants, F80I/F83I and F83I/F80I, showed a reduced (~ 2.5-fold) channel conductance as compared to WT [97]. These results suggest that each Fluc homodimer forms two functionally active pores.

One of the main components of the pore structure is TM4, whose pore-facing surface is lined with conserved H-bonding side-chains that form a 'polar-track' on each side of the central Phe residues (Figure 3.5). In Fluc-Ec2, several residues forming the polar-track were mutated to Ala to test the dependence of the proposed permeation pathway on polarity. Unexpectedly, S84A, S102A, and S110A mutants left channel activity unaltered. On the contrary, H106A and T114A mutants, as well as alternate H-bonding mutants such as H106S, H106Y, H106W, H106N, H106F, N41S, N41Q, and N41Y, showed no conductance at all. The residue T114 deserves special attention since, even though lying far from the electron densities interpreted as F<sup>-</sup>, its substitution by the hydrophobic Ala results in an inactive channel. Surprisingly, the polar substitute T114S also abolishes channel activity, whereas T114V and T114I mutants retained full activity. These results leave the mysterious role of T114 unresolved. Further tests were carried out in which the conserved Phe residues were mutated with aromatic substitutions (Y, W, H), aliphatic H-bonding residues (S, T, Q), hydrophobic aliphatics (L, A), as well as a non-aromatic F<sup>-</sup>-coordination substitute (M). All mutations, save for F80M, rendered non-functional channels [98].

All these results highlight the relevance of several residues for the function of Fluc. However, they do not elucidate the detailed permeation and selectivity mechanisms.

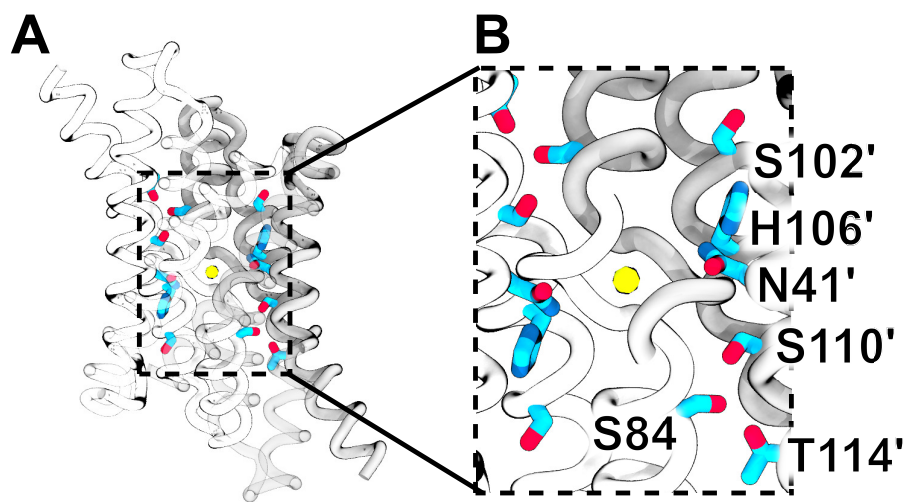


Figure 3.5: **Residues of Fluc-Ec2 polar-track.** (A) Crystal structure of Fluc-Ec2. (B) Close-up of the residues forming the polar-track in Fluc-Ec2.

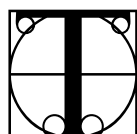




---

## MOTIVATION

---



THIS THESIS aims to study cell membrane channels by means of computational methods. At the beginning of this project the aim was to build a membrane channel database based on channel permeabilities. For that purpose, we needed a fast and accurate computational method able to compute the Potential of Mean Force (PMF) of solute permeation across membrane channels. The 3-Dimensional Reference Interaction-Site Model (3D-RISM) was presented as a fast and accurate method for computing PMFs for the permeation of different solutes across aquaporin channels by Phongphanphane *et al.* (2010) [99]. However, we compared the PMFs calculated with 3D-RISM to the PMFs calculated with Molecular Dynamics (MD) coupled with Umbrella Sampling (US), which is a widely and commonly used computational method to calculate PMFs, and concluded that 3D-RISM is not suitable to compute PMFs of solute permeation across protein channels [65].

We shifted our attention to Fluc-Bpe channel (see Chapter 3), whose novel crystal structure was analyzed by means of computational methods, such as MD, US, Thermodynamic Integration (TI), and Quantum Mechanics (QM) simulations [91]. In addition, we aimed to shed light onto the unanswered questions about Fluc: i) what is the permeation mechanism, and ii) what is the mechanism for the selectivity of  $F^-$  over  $Cl^-$ .



Part II

METHODS

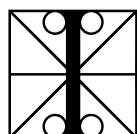


# 5

---

## COMPUTATIONAL THEORY AND METHODS

---



IN THIS CHAPTER the general theory of the methods used in this thesis will be described. The included methods are: Molecular Dynamics (MD) simulations, Potential of Mean Force (PMF) calculations via Umbrella Sampling (US), free energy calculations by Thermodynamic Integration (TI), Quantum-chemical calculations, and PMF calculations via the 3-Dimensional Reference Interaction-Site Model (3D-RISM).

### 5.1 MOLECULAR DYNAMICS

As outlined in Section 1.3, MD simulations provide an invaluable tool to study the thermodynamic properties of biological macromolecules [100]. In short, MD simulations employ three different approximations: i) the decoupling of the motion of nuclei and electrons (Born-Oppenheimer approximation), ii) the classical description of the motion of the nuclei, and iii) the approximation of the potential energy surface by an empirical force-field [101]. In this thesis, most of the calculations have been carried out using classical MD with empirical force-fields.

#### 5.1.1 Born-Oppenheimer approximation

The separation of the fast electronic motions and the nuclear motion allows the description of a molecular system as a set of nuclei moving over a potential energy surface provided by the electrons. The time-evolution of such a molecular system can be, thus, described by the time-dependent Schrödinger equation [102]:

$$H\psi = i\hbar \frac{\partial \psi}{\partial t} \quad (5.1)$$

where  $H$  is the Hamiltonian of the system,  $\hbar$  is the Planck constant  $h$  divided by  $2\pi$ , and  $\psi$  is the wave function which depends both on

the positions of the nuclei  $\mathbf{R}$  and the electrons  $\mathbf{r}$ , and the time  $t$  such that:

$$\psi = \psi(\mathbf{R}, \mathbf{r}, t) \quad (5.2)$$

$$= \psi_n(\mathbf{R}, t)\psi_e(\mathbf{r}; \mathbf{R}) \quad (5.3)$$

Here  $\psi(\mathbf{R}, \mathbf{r}, t)$  is split into the wave function of the nuclei  $\psi_n(\mathbf{R}, t)$  and the wave function of the electrons  $\psi_e(\mathbf{r}; \mathbf{R})$ , which is parametrically dependent on the nuclei positions. Hence, with fixed nuclei positions,  $\psi_e(\mathbf{r}; \mathbf{R})$  can be derived from the time-independent Schrödinger equation that contains the nuclei positions  $\mathbf{R}$  only parametrically:

$$H_e(\mathbf{R})\psi_e(\mathbf{r}; \mathbf{R}) = E_e(\mathbf{R})\psi_e(\mathbf{r}; \mathbf{R}) \quad (5.4)$$

where the electronic Hamiltonian  $H_e(\mathbf{R}) = H - \Gamma_n$  is equivalent to the subtraction of the kinetic energy of the nuclei  $\Gamma_n$  from the Hamiltonian of the system  $H$ .  $E_e(\mathbf{R})$  is the adiabatic or Born-Oppenheimer potential-energy surface, which can be expressed as a diagonal matrix with the eigenvalues of Equation 5.4. Now the time-evolution of the nuclei can be defined as a time-dependent Schrödinger equation:

$$(\Gamma_n + E_e(\mathbf{R}))\psi_n(\mathbf{R}, t) = i\hbar \frac{\partial \psi_n(\mathbf{R}, t)}{\partial t} \quad (5.5)$$

which holds as long as the potential energy surfaces ( $E_e(\mathbf{R})$ ) of different excited states differ from each other, as it usually is for molecules in the ground state.

### 5.1.2 Classical MD approximation

In the classical approximation, nuclei are represented as classical particles whose time-evolving trajectories follow Newton's second law:

$$m_i \frac{\partial^2 \mathbf{R}_i}{\partial t^2} = -\nabla_{\mathbf{R}_i} V(\mathbf{R}) \quad (5.6)$$

$$m_i \mathbf{a}_i = \mathbf{F}_i \quad (5.7)$$

where  $m_i$ ,  $\mathbf{R}_i$ ,  $\mathbf{a}_i$ , and  $\mathbf{F}_i$  denote the mass, position, acceleration, and force on atom  $i = 1, \dots, n$  in a system of  $n$  atoms, respectively, and  $V(\mathbf{R})$  is the potential energy surface of the system. The classical approximation is, in general, appropriate to study of protein dynamics under non-extreme conditions, which is usually the case. In solvent, we observe proteins as classical actors due to the influence of decoherence [103, 104]. However, many important processes involving charge transfer, the formation or breaking of covalent bonds, and excited electron states, are left out the scope of the classical approximation.

## 5.1.3 Empirical Force-Fields

## 5.1.3.1 Additive Force-Fields

An additive empirical force-field is the sum of simple empirical expressions that yield a sufficiently accurate approximation to the potential energy surface while being computationally inexpensive. Accordingly, the molecules are represented by a ball-and-spring-like model, in which atoms are spheres with a single point partial charge. A typical functional form of such a force-field includes the sum of bonded and non-bonded interactions between the particles in the system [42, 105]:

$$\begin{aligned}
 V(\mathbf{R}) &= \overbrace{V_{\text{bonds}} + V_{\text{angles}} + V_{\text{dihedrals}} + V_{\text{imp. dih.}}}^{\text{bonded}} + \overbrace{V_{\text{LJ}} + V_{\text{coulomb}}}^{\text{non-bonded}} \\
 &= \sum_{\text{bonds } i} \frac{k_i}{2} (l_i - l_{i,0})^2 \\
 &\quad + \sum_{\text{angles } i} \frac{f_i}{2} (\varphi_i - \varphi_{i,0})^2 \\
 &\quad + \sum_{\text{dihedrals } i} \frac{V_i}{2} [1 + \cos(n\varphi_i - \varphi_{i,0})] \\
 &\quad + \sum_{\text{imp. dih. } i} \kappa_i (\xi_i - \xi_{i,0})^2 \\
 &\quad + \sum_{\text{pairs } i,j} 4\epsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{4\pi\epsilon_0\epsilon_r r_{ij}}
 \end{aligned} \tag{5.8}$$

Here, the bond potential  $V_{\text{bonds}}$ , the bond angle potential  $V_{\text{angles}}$ , and the improper dihedral potential  $V_{\text{imp. dih}}$ —which describe out-of-plane bending modes—are modeled as harmonic potentials, whereas the proper dihedral potential  $V_{\text{dihedrals}}$  is modeled by a cosine with periodicity  $n$  and potential barriers  $V_i$ . The non-bonded interactions, which are evaluated pairwise, include the short-range repulsive and attractive dispersion interactions generally represented by a Lennard-Jones (LJ) potential, in which the parameters  $\sigma_{ij}$  and  $\epsilon_{ij}$  define the width and strength of the potential. In the Coulombic term that models the electrostatic interactions,  $q_i$  denotes the partial charge of atom  $i$ , and the relative dielectric constant  $\epsilon_r$  is typically set to 1.

There is a wide variety of force-fields that are commonly used for biomolecular simulations, which include protein, nucleic acid, lipid, and sugar-like molecule simulations, e.g. AMBER [106–116], CHARMM [117–127], GROMOS [128–132], and OPLS [133–136]. Additional force-fields are used to parametrize “non-standard” small molecules such as drugs and ligands. Among them, GAFF [137] and CGenFF [138] are the most common force-fields.

### 5.1.3.2 Polarizable Force-Fields

The accuracy in the representation of many biological and chemical condensed-phase systems is limited by the fixed partial atomic charge convention of additive force-fields. The electronic distribution of molecules changes in reaction to the local electric field, i. e. they are polarizable, but this is not reflected in the above-mentioned empirical additive force-fields. To overcome this limitation, an extended definition of force-fields that allow explicit electronic polarization is desirable [31]. Many methods have been developed to include explicit polarization in MD simulations [139, 140], among which the three main approaches are the point-dipole [141–143], the fluctuating charge [144–146], and the Drude oscillator model [31, 147, 148]. In this section, we will focus on the latter, since it was used in some simulations described here.

The Drude oscillator model is an extension of empirical additive force-fields, in which many of the pairwise features are kept in the functional form. The extension is based on the attachment of an auxiliary charged particle to all non-hydrogen atoms. This particle, which is attached via a harmonic spring, reacts to the induced polarization effect under the influence of an external electric field. The model also includes, usually in hydrogen bond-acceptor atoms, electron lone pairs represented with a virtual site. The charge  $q$  of a given atom is divided into the core atom charge  $q_A$  and the charge in the Drude oscillator  $q_D$  such that  $q = q_A + q_D$ . Therefore, the isotropic atomic polarizability is given by

$$\alpha = \frac{q_D^2}{k_D} \quad (5.9)$$

where  $k_D$  is the force constant of the atom-Drude harmonic bond. For a given  $\alpha$ ,  $k_D$  can be defined such that the Drude particle oscillates around a displaced position  $\mathbf{r} + \mathbf{d}$  in the presence of a uniform electric field  $\mathbf{E}$ . The displacement  $\mathbf{d}$  of the Drude particle is given by

$$\mathbf{d} = \frac{q_D \mathbf{E}}{k_D} \quad (5.10)$$

from which the induced atomic dipole  $\mu$  is derived as

$$\mu = \frac{q_D^2 \mathbf{E}}{k_D} \quad (5.11)$$

Hence, adding polarizable Drude particles to a system of atoms yields the following functional form, very similar to Equation 5.8

$$\begin{aligned} V(\mathbf{R}, \mathbf{d}) = & V_{\text{self}}(\mathbf{d}) \\ & + V_{\text{bonds}}(\mathbf{R}) + V_{\text{angles}}(\mathbf{R}) + V_{\text{dihedrals}}(\mathbf{R}) + V_{\text{imp. dih.}}(\mathbf{R}) \\ & + V_{\text{LJ}}(\mathbf{R}) + V_{\text{coulomb}}(\mathbf{R}, \mathbf{d}) \end{aligned}$$



$$(5.12)$$

where  $V_{\text{self}}(\mathbf{d})$  is the harmonic self-polarization term, which is expressed as

$$V_{\text{self}}(\mathbf{d}) = \frac{1}{2}k_{\text{D}}\mathbf{d}^2 \quad (5.13)$$

in the case of isotropic polarization. For anisotropic polarization, i. e. when the spring constant of the Drude oscillator is different in different directions,  $V_{\text{self}}$  is defined as

$$V_{\text{self}} = \frac{1}{2}([K_{11}^{(\text{D})}]d_1^2 + [K_{22}^{(\text{D})}]d_2^2 + [K_{33}^{(\text{D})}]d_3^2) \quad (5.14)$$

where the isotropic force constant  $k_{\text{D}}$  is expanded to a tensor,  $\mathbf{K}^{\text{D}}$ , with off-diagonal elements set to zero, and  $d_1$ ,  $d_2$ , and  $d_3$  represent the components of the Drude-atom displacement vector  $\mathbf{d}$  defined with a coordinate frame based on the local atomic geometry. For example  $d_1 = \mathbf{d} \cdot \hat{\mathbf{n}}_{i,j}$  where  $\hat{\mathbf{n}}_{i,j}$  is a unit vector directed between atoms  $i$  and  $j$  in the molecule of interest. The term  $V_{\text{coulomb}}(\mathbf{R}, \mathbf{d})$  from Equation 5.12 represents all Coulombic interactions i. e. atom-atom, atom-Drude, and Drude-Drude.

The last remarkable difference of the Drude oscillator model compared with common additive force-fields is the treatment of dipole-dipole interactions for atoms connected by one or two bonds —also known as 1-2 and 1-3 interactions—. In additive force-fields, these interactions are implicitly represented in the parametrization of  $V_{\text{bonds}}$  and  $V_{\text{angles}}$  (see Equation 5.8), whereas in the Drude oscillator model they are explicitly included, and screened by

$$S_{ij}(r_{ij}) = 1 - \left[ \left( 1 + \frac{ar_{ij}}{2(\alpha_i\alpha_j)^{1/6}} \right) \right] \exp \left[ \frac{-ar_{ij}}{(\alpha_i\alpha_j)^{1/6}} \right] \quad (5.15)$$

where  $r_{ij}$  is the distance between atoms  $i$  and  $j$ ,  $\alpha_i$ ,  $\alpha_j$  are the respective atomic polarizabilities, and  $a$  is a damping constant. Equation 5.15 is equivalent to the smeared charge distribution proposed by Thole [149].

#### 5.1.4 MD algorithms

In this thesis, all the MD simulations are carried out using the Gromacs software package [35, 150–153]. In the following paragraphs, several relevant details of the MD algorithms used in Gromacs are discussed.

**TIME INTEGRATION** In Gromacs, the time integration is commonly carried out via the *leap-frog* algorithm [154], by which velocities  $\mathbf{v}$  and positions  $\mathbf{r}$  of atoms are updated as

$$\mathbf{v}(t + \frac{1}{2}\Delta t) = \mathbf{v}(t - \frac{1}{2}\Delta t) + \frac{\Delta t}{m}\mathbf{F}(t) \quad (5.16)$$

$$\mathbf{r}(t + \Delta t) = \mathbf{r}(t) + \mathbf{v}(t + \frac{1}{2}\Delta t)\Delta t \quad (5.17)$$

where the time-step  $\Delta t$  must be significantly smaller than the fastest period of the motion in the system—the vibrations of hydrogen atoms—to avoid integration errors. While this would require a time-step of 1 fs, fortunately not all vibrational modes need to be explicitly simulated. Bond lengths can be constrained, thereby allowing the use of a 2 fs time-step. Typically, water molecule bonds are constrained with the SETTLE [155] algorithm, whereas the remaining bonds are constrained using the LINCS [156, 157] algorithm. Hydrogen atoms are modeled as virtual-sites [158] in many of the simulations performed in this thesis, allowing a time-step of 4 fs [153].

**LENNARD-JONES AND ELECTROSTATIC INTERACTIONS** If interactions of all pairs of particles would be computed, the calculation of the pairwise long-range non-bonded interactions would scale quadratically with the number of atoms  $N$  in a system. To prevent such a disadvantageous scaling, LJ interactions are cut off between 1.0–1.4 nm [159]. The most time-demanding part in a simulation is the calculation of electrostatic interactions. Unfortunately, a cut-off in the Coulomb potential induces serious artifacts [160, 161] due to its low decay with  $r^{-1}$ . Therefore, long-range coulomb interactions are typically treated via the Particle-Mesh Ewald (PME) method [162, 163], leading to a more affordable scaling of  $N \log N$  instead of  $N^2$ .

**TEMPERATURE AND PRESSURE COUPLING** Biomolecular MD simulations are usually carried out in the NPT ensemble, i. e. at isobaric-isothermal conditions. Therefore, the simulation system is coupled to i) a temperature bath, to control temperature and to avoid the heating of the system due to numerical imprecisions and cut-offs, and ii) a pressure bath, which prevents from the high pressure that might arise due to the low compressibility of the condensed media in a fixed volume. The velocity-rescale scheme [164] was used for temperature coupling in most of the simulations in this thesis, whereas the Berendsen barostat [165] and the Parrinello-Rahman [166, 167] barostat were used for pressure coupling in equilibration and production simulations, respectively.

**THE SCF REGIME AND THE LAGRANGIAN FRAMEWORK** In polarizable MD simulations, the efficiency of the integration scheme is a major issue as the calculation of all the induced dipoles is computa-

tionally very demanding. Two main methods can be used to perform simulations with Drude oscillator model. The first one is the Self Consistent Field (SCF) method, an extension of the Born-Oppenheimer approximation, in which the positions of the Drude oscillators are energy minimized at each integration step such that

$$\frac{\partial V}{\partial \mathbf{d}_i} = 0 \quad (5.18)$$

where the index  $i$  runs over all the Drude-atom pairs. This method not always converges given a limited numerical precision, plus it can accumulate numerical errors due to different initial guesses for the induced dipoles, or minimization algorithms [168]. The most common alternative to the SCF method is the Lagrangian framework [148], in which an effective mass  $m_D$  is attributed to each Drude oscillator, such that the positions of the Drude oscillators can be likewise integrated with the atomic positions of the system. The convention is to set  $m_D = 0.4$  amu, which is subtracted from the core atom such that the total mass of the Drude-atom pair remains equal to the atomic mass.

## 5.2 POTENTIAL OF MEAN FORCE

Originally introduced by Kirkwood (1935) [169], the Potential of Mean Force (PMF) is an invaluable concept to investigate multi-dimensional data, such as the high-dimensional free-energy landscape of biomolecular processes. The PMF  $W(\xi)$  is the resulting potential from integrating the mean force acting on the particle(s) along the reaction coordinate  $\xi$ . It is defined from the average distribution function  $\langle p(\xi) \rangle$  as [170]:

$$W(\xi) = W(\xi^*) - k_B T \ln \left[ \frac{\langle p(\xi) \rangle}{\langle p(\xi^*) \rangle} \right] \quad (5.19)$$

$$\langle p(\xi) \rangle = \frac{\int d\mathbf{R} \delta(\xi'[\mathbf{R}] - \xi) \exp[-U(\mathbf{R})/k_B T]}{\int d\mathbf{R} \exp[-U(\mathbf{R})/k_B T]} \quad (5.20)$$

where  $k_B T$  is the product of the Boltzmann constant and the temperature,  $\xi^*$  is an arbitrary reference,  $\langle \bullet \rangle$  is the ensemble average,  $U(\mathbf{R})$  is the potential energy of the system as a function of its coordinates  $\mathbf{R}$ , and  $\xi'[\mathbf{R}]$  is the reaction coordinate mapping function that usually depends on a few degrees of freedom. More intuitively, the PMF  $W(\xi)$  is exactly related to a projection [171]: consider a system of  $N$  particles with  $3N$  spacial coordinates in which a particle moves in an external potential  $V(x_1, \dots, x_{3N})$ , so that  $p(x_1, \dots, x_{3N}) \propto \exp[-V(x_1, \dots, x_{3N})/k_B T]$  is the Boltzmann distribution. Then we can write PMF  $W(x_1)$  as

$$W(x_1) = -k_B T \ln \int dx_2, \dots, dx_{3N} \exp[-V(x_1, \dots, x_{3N})/k_B T]. \quad (5.21)$$

*In the entry "What I have against (most) PMF calculations" from the Statistical Biophysics Blog, Zuckerman discusses on the most common pitfalls in calculating and interpreting PMFs [173]*

Another key idea is that the PMF  $W(\xi)$  is a *Gibbs* free energy, i. e.  $G(\xi)$ , the proof of which is given in reference [172]. Although PMFs can be computed in many different ways, the choice of the reaction coordinate(s), along which the sampling will take place, can be very challenging. The difficulty of choosing the proper reaction coordinate(s) has fueled the development of many methods addressed to solve that problem [174–176]. In this thesis, PMFs will be calculated via US simulations, which will be introduced in the next section.

### 5.2.1 Umbrella Sampling

Umbrella Sampling (US) is an enhanced sampling method used to aid in the sampling of functionally relevant events, such as the solute permeation across a membrane channel, which are typically difficult to observe due to the limited simulation time in equilibrium MD simulations. First proposed by Torrie & Valleau (1974), US divides the reaction coordinate  $\xi$  in multiple *windows*. A harmonic biasing potential is applied to each of them to ensure sampling along the whole  $\xi$  [177]. Hence, the biasing potential  $\omega_i(\xi)$  with a force constant  $K$  of window  $i$ , which is centered at  $\xi_i$ , is defined as [170, 178]:

$$\omega_i(\xi) = \frac{1}{2}K(\xi - \xi_i)^2 \quad (5.22)$$

An independent biased simulation is carried out for each window  $i$ , from which the unbiased PMF  $G_i(\xi)$  is recovered following

$$G_i(\xi) = -k_B T \ln P_i^b(\xi) - \omega_i(\xi) + F_i \quad (5.23)$$

$$F_i = -k_B T \ln \langle \exp[-\omega_i(\xi)/k_B T] \rangle \quad (5.24)$$

where  $P_i^b$  denotes the biased distribution of window  $i$ , which is extracted from a biased MD simulation, and  $F_i$  the free energy associated with introducing the umbrella potential [170]. There are several ways to calculate the unbiased PMF  $G(\xi)$  from biased US [170, 179, 180]. In this thesis, the Weighted Histogram Analysis Method (WHAM) [179] will be used as implemented in the wham module of Gromacs [181].

## 5.3 THERMODYNAMIC INTEGRATION

Thermodynamic Integration (TI) is a method used to calculate free energy differences in which a parameter specifying the thermodynamic state is slowly varied. In many cases, the underlying physical process we want to study, such as solvation or ligand binding, can be defined as a composition of non-physical transformations, which turn out to be much easier to simulate via thermodynamic cycle than the physical process itself [182–187]. TI is very common nowadays and has been successfully used for various purposes, such as rational drug design

[188], standard binding free energies [189], or solvation free energies [182]. In the isobaric-isothermal NPT ensemble, the *Gibbs* free energy difference  $\Delta G_{AB}$  between the thermodynamic states A and B can be defined as:

$$\Delta G_{AB} = G_B - G_A = -k_B T \ln \frac{Z_B}{Z_A} \quad (5.25)$$

where  $Z$  is the partition function in the NPT ensemble:

$$Z(N, P, T) = \frac{1}{h^{3N} N!} \int dV \iint d\mathbf{p} d\mathbf{r} \exp \left[ \frac{H(\mathbf{p}, \mathbf{r}) + PV}{k_B T} \right]. \quad (5.26)$$

Here  $N$ ,  $P$ ,  $T$ , and  $V$  are the number of particles, the pressure, the temperature, and the volume, respectively.  $h$  represents the Planck constant, and the Hamiltonian  $H(\mathbf{p}, \mathbf{r})$  gives the total energy of the system for a given set of momenta  $\mathbf{p}$  and coordinates  $\mathbf{r}$ . The Hamiltonian can be, in turn, split into a kinetic  $K(\mathbf{p})$  and potential energy term  $U(\mathbf{r})$ :

$$H(\mathbf{p}, \mathbf{r}) = K(\mathbf{p}) + U(\mathbf{r}) \quad (5.27)$$

$$K(\mathbf{p}) = \sum_{i=1}^N \frac{\mathbf{p}_i^2}{2m_i} \quad (5.28)$$

where  $m_i$  is the mass of particle  $i$ . If the masses in the states A and B are equal, the kinetic term cancels out [187]. The non-physical (alchemical) path connecting the states A and B is described by a variable  $\lambda$ , where  $\lambda = 0$  and  $\lambda = 1$  correspond to the states A and B, respectively. In this thesis, TI was used such that each  $\lambda$ -step was performed in a separate simulation, in which the partial derivative is calculated analytically, thus, the free energy difference  $\Delta G$  can be calculated as the integral over a generalized mean force on  $\lambda$  [186, 187]:

$$\Delta G_{AB} = \int_0^1 d\lambda \left\langle \frac{\partial U(\lambda, (\mathbf{r}))}{\partial \lambda} \right\rangle_{\lambda}. \quad (5.29)$$

#### 5.4 3-DIMENSIONAL REFERENCE INTERACTION-SITE MODEL

3D-RISM is based on the integral equation theory of molecular liquids. It allows evaluating the thermodynamic properties of a system by computing solvent density distributions around a complex solute on a three-dimensional grid [190–192]. In this section, we will use “solute” and “solvent” as it is used in the 3D-RISM literature. Accordingly, in the context of small-molecule permeation across a protein channel, “solute” refers to the protein, and “solvent” to bulk water *and*, if present, the additional permeating small molecules. Note the contrast to the nomenclature adopted in literature on protein channels, in which the term “solute” refers to the permeating small molecule.

${}_3\text{D-RISM}$  integral equation is derived from the six-dimensional molecular Ornstein-Zernike (OZ) equation for a solute-solvent mixture of rigid molecules at infinite dilution. The OZ equation is defined as [192, 193]

$$h^{uv}(1,2) = c^{uv}(1,2) + \int c^{uv}(1,3)\rho^v h^{vv}(3,2)d(3), \quad (5.30)$$

where the superscripts “u” and “v” denote solute and solvent species, respectively,  $h^{uv}(1,2) = h^{uv}(r_{12}, \Omega_1, \Omega_2)$  is the solute-solvent total correlation function,  $c^{uv}(1,2) = c^{uv}(r_{12}, \Omega_1, \Omega_2)$  is the solute-solvent direct correlation function, and  $h^{vv}(3,2) = h^{vv}(r_{32}, \Omega_3, \Omega_2)$  is the solvent-solvent total correlation function.  $h^{uv}$ ,  $c^{uv}$ , and  $h^{vv}$  depend on the separation  $r_{12}, r_{32}$  of molecules of the corresponding species, as well as on the molecule orientations  $\Omega_1, \Omega_2$ , and  $\Omega_3$ .  $\rho^v$  denotes the solvent number density. The integral is taken over  $d(3) = dr_3 d\Omega_3$ , where  $r_3$  is the position of a third molecule.

The standard site-site dielectrically consistent RISM equation (DRISM) [194, 195] carries out the full orientational reduction of Equation 5.30. This yields the spatial correlation functions between the solvent atomic species sites, which are used subsequently to solve the  ${}_3\text{D-RISM}$  equation. In turn,  ${}_3\text{D-RISM}$  averages out the solvent molecular orientations, but keeps the orientational description of the solute molecule [190]:

$$h_{\alpha}^{uv}(\mathbf{r}) = \sum_{\gamma=1}^N \int_{\mathbb{R}^3} c_{\gamma}^{vu}(\mathbf{r}-\mathbf{r}') \chi_{\gamma\alpha}^{vv}(|\mathbf{r}'|) d\mathbf{r}', \quad (5.31)$$

where  $\alpha = 1, \dots, N$ ,  $\chi_{\gamma\alpha}(\mathbf{r})$  is the bulk solvent susceptibility function taken from DRISM,  $\gamma$  and  $\alpha$  denote the indexes over the atomic species sites in a solvent molecule, and  $N$  is the number of atomic species sites in the solvent molecule.

Solving Equation 5.30 requires multiple integrals that are difficult to solve except for some special cases. Therefore, to solve the OZ integral equation, both DRISM and  ${}_3\text{D-RISM}$  have to be complemented by a closure relation. There are different closure relations such as the Percus-Yevick (PY), HyperNetted-Chain (HNC) or the Mean-Spherical Approximation (MSA) [193, 196, 197], by which the solution to Equation 5.30 can be approximated. In this thesis, we mainly use the Kovalenko-Hirata (KH) closure relation, as implemented in AMBER,[198] which combines the HNC closure for low-density regions with the MSA closure for high density regions [192]:

$$h_{\alpha}^{uv}(\mathbf{r}) = \begin{cases} \exp [d_{\alpha}^{uv}(\mathbf{r})] - 1 & \text{for } d_{\alpha}^{uv}(\mathbf{r}) \leq 0 \\ d_{\alpha}^{uv}(\mathbf{r}) & \text{for } d_{\alpha}^{uv}(\mathbf{r}) > 0 \end{cases}, \quad (5.32)$$

where

$$d_{\alpha}^{uv}(\mathbf{r}) = -\frac{1}{k_B T} u_{\alpha}^{uv}(\mathbf{r}) + h_{\alpha}^{uv}(\mathbf{r}) - c_{\alpha}^{uv}(\mathbf{r}). \quad (5.33)$$

Here,  $u_{\alpha}^{\text{uv}}(\mathbf{r})$  denotes the interaction potential between the solute and solvent site  $\alpha$ , given by the sum of the electrostatic and LJ potentials between the solvent site  $\alpha$  and all solute atoms.

#### 5.4.1 3D-RISM solution can be translated into a PMF

The approximated solution to the 3D-RISM integral equation can be translated into a solute-solvent three-dimensional density distribution function [190]

$$g_{\alpha}^{\text{uv}}(\mathbf{r}) = h_{\alpha}^{\text{uv}}(\mathbf{r}) + 1 \quad (5.34)$$

where  $\alpha$  denotes the atomic species site of the solvent permeating molecule.  $g_{\alpha}^{\text{uv}}(\mathbf{r})$  can be, in turn, reduced to the one-dimensional density distribution function  $g_{\alpha}^{\text{uv}}(z)$  by integrating out the  $x$  and  $y$  coordinates:

$$g_{\alpha}^{\text{uv}}(z) = \int_{A_g(z)} g_{\alpha}^{\text{uv}}(\mathbf{r}) \, dx dy. \quad (5.35)$$

where  $A_g(z)$  denotes the integration area. Hence, from the one-dimensional density distribution function, the calculation of the corresponding PMF is straightforward:

$$\Delta G_{\alpha}(z) = -k_B T \ln \frac{g_{\alpha}^{\text{uv}}(z)}{g_{\alpha}^{\text{uv}}(z_{\text{bulk}})}, \quad (5.36)$$

where  $g_{\alpha}^{\text{uv}}(z_{\text{bulk}})$  is calculated similarly to  $g_{\alpha}^{\text{uv}}(z)$ , and  $z_{\text{bulk}}$  is a  $z$ -position in bulk sufficiently far from the protein.





---

## COMPUTATIONAL DETAILS

---

### 6.1 SYSTEM SET-UP FOR US SIMULATIONS OF UT-B

Initial coordinates of UT-B were taken from the Protein Data Bank (PDB) ([www.rcsb.org](http://www.rcsb.org)) [199] (PDB ID: 4EZC) [64]. Initial coordinates for the lipid bilayer were taken from a 1 ns equilibrium simulation of a 328 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine (POPC) lipid hydrated membrane. The protein was embedded in the bilayer with the `g_membed` software [45], and the simulation box was filled by explicit water molecules and neutralized by one counter ion. The final simulation system contained the UT-B trimer, 274 POPC molecules, 24989 TIP3P [200] water molecules, and 1 Na<sup>+</sup> ion.

Protein atom interactions were described by the Amber ff99SB\*-ILDN force-field [109], and lipid parameters were taken from Berger *et al.* [201, 202]. The simulations were performed with the Gromacs simulation software (version 4.6) [152]. Electrostatic interactions were calculated at every step with the PME method [163]. Short-range repulsive and attractive dispersion interactions were described by a LJ potential with a cut-off at 1 nm. The geometry of water molecules was constrained with the SETTLE algorithm [155], and all other bond lengths were constrained with LINCS [157]. Hydrogen atoms of the protein were constructed as virtual sites, allowing a 4 fs time step [158]. The simulation temperature was kept constant at 300 K using velocity rescaling ( $\tau = 2.5$  ps) [164], and the pressure was kept at 1 bar with a semiisotropic Berendsen barostat ( $\tau = 2$  ps) [165]. The system was equilibrated for 150 ns before production. Figure 6.1 shows a typical simulation box.

PMFs for permeation across UT-B were computed for the following solutes: TIP3P water [200], ammonia (NH<sub>3</sub>), urea, molecular oxygen (O<sub>2</sub>), and methanol. Parameters for NH<sub>3</sub> were generated using AmberTools15 from the AMBER molecular dynamics software package [203], with AM1-BCC charges [204, 205], and the atoms were defined with the OPLS-AA force-field [133, 135]. Urea parameters were taken from Duffy *et al.* [206]. LJ parameters for O<sub>2</sub> were taken from the CHARMM22 force-field [117] and partial charges from Hub & de Groot [207]. For methanol, we used the GAFF parameters provided at <http://virtualchemistry.org/> [208, 209].

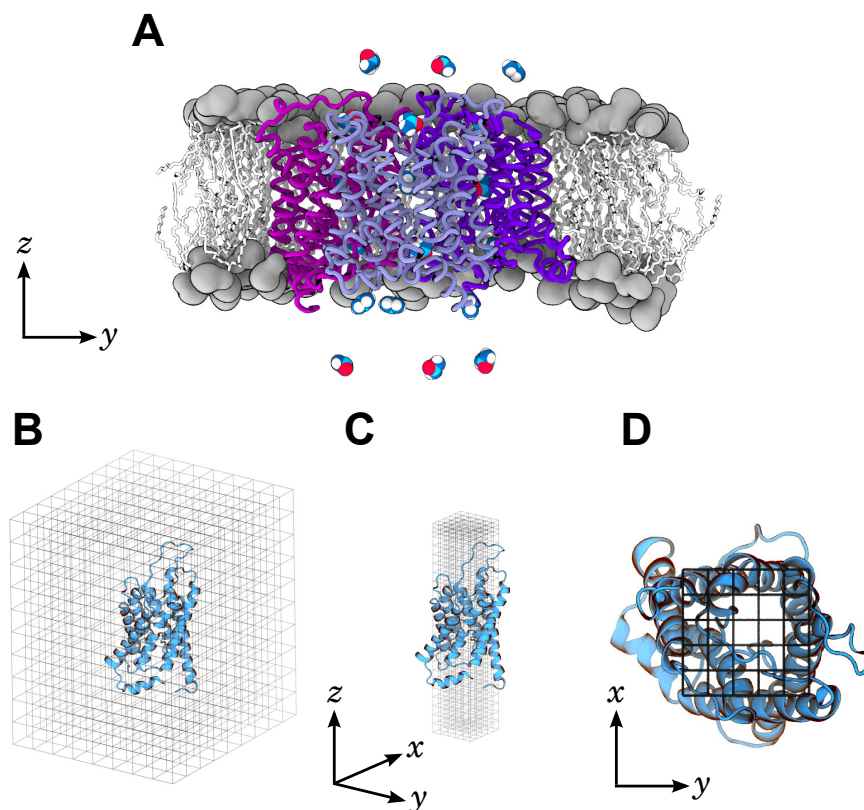


Figure 6.1: **US simulation system box with UT-B and 3D-RISM calculation grid.** (A) Three-dimensional representation of a typical US simulation system box. The lipid heads are represented as grey volume, and the lipid tails are represented as white sticks. The UT-B protein is shown in purple, blue, or dark blue ribbon representation. The urea molecules placed along the reaction coordinate  $z$  (membrane normal) are depicted as coloured spheres. (B) Schematic view of AQP1 inside a 3D-RISM grid. (C) Reduced three dimensional grid across the AQP1 pore. (D)  $x$  and  $y$  view of the aforementioned reduced area. The grids do not show the actual number of grid points.

### 6.1.1 US simulation details

The  $z$ -coordinate (membrane normal) was taken as reaction coordinate, measured as the distance between the center-of-mass (COM) of the solute molecule to COM of the transmembrane backbone atoms of the protein. Starting frames for US were taken from a 150 ns equilibrium simulation. 21 solute molecules per simulation (7 per monomer) were placed, allowing the collection of 21 umbrella windows from each simulation. The solutes were separated by 1.5 nm in the  $z$  direction and restrained by a harmonic umbrella potential with a force constant of  $1000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$ . In addition, to ensure that the solutes adopt a well-defined membrane area outside the channel (in bulk), the solutes were restrained into cylinders aligned along the respec-

tive channel axis, as suggested by Allen *et al.* [210]. In this work, the cylindrical restraint was implemented as a flat-bottomed quadratic potential,

$$V_{\text{cyl}}(r) = \frac{1}{2}k_c(r - r_c)^2H(r - r_c), \quad (6.1)$$

in which  $r$  represents the distance from the cylinder axis that was aligned along the respective pore,  $k_c = 1000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$  is the force constant,  $r_c = 0.8 \text{ nm}$  is the cylinder radius, and  $H$  is the Heaviside step function. All the US simulations were carried out using an in-house modified Gromacs simulation software (version 4.6) which implemented the flat-bottomed quadratic potential.

Adjacent umbrella windows were separated by 0.01 nm in simulations of water, urea, and  $\text{NH}_3$ , hence, 150 simulation systems were required to cover the complete reaction coordinate. For  $\text{O}_2$  and methanol, adjacent windows were separated by 0.025 nm, thus, 60 simulation systems were required. US simulations for urea were conducted for 10 ns, and all other US simulations for 5 ns. The temperature was kept constant using a stochastic dynamics integration scheme ( $\tau = 0.5 \text{ ps}$ ) [211], and the pressure was kept at 1 bar with a semiisotropic Parrinello-Rahman barostat ( $\tau = 5 \text{ ps}$ ) [166, 167]. A 2 fs time step was used in the US simulations. All other simulation parameters were identical to the equilibrium simulations.

### 6.1.2 PMF construction after US

We removed the first 2 ns of each trajectory for equilibration in simulations with water,  $\text{NH}_3$ ,  $\text{O}_2$ , and methanol, and the first 4 ns in simulations with urea. In total, 11,970 umbrella histograms were collected from 3.6  $\mu\text{s}$  of simulation of the UT-B system. PMFs were calculated using a periodic implementation of WHAM [179, 181]. We collected histograms from the three channels for each solute. Then, the histograms were combined into one PMF. Since we applied a cylindrical flat-bottomed potential, the PMFs refer to a channel density of one channel per cross-section area of the cylinder. Here, we corrected the PMFs such that they correspond to a channel density of one channel per membrane area occupied by a UT-B monomer. Hence, we used a trapezoidal correction between the entrance and exit regions of the pore as done previously[207, 212], shifting the PMF by

$$k_B T \ln \left( \frac{A_{\text{mono}}}{A_{\text{cyl}}} \right), \quad (6.2)$$

where  $A_{\text{mono}}$  is the area occupied by a UT-B monomer and  $A_{\text{cyl}}$  is the effective cross-section area of the cylinder. The area of the cylinder was approximated as

$$A_{\text{cyl}} = \pi(r_c + 2\sigma_c)^2 \quad (6.3)$$

$$\sigma_c = \sqrt{\left(\frac{k_B T}{k_c}\right)} \quad (6.4)$$

where  $\sigma_c$  is the width of the Gaussian-shaped distribution in the quadratic region of the flat-bottomed cylindrical potential. In fact, the entropy of a uniform 2-dimensional distribution in  $A_{\text{cyl}}$  approximately equals the entropy of a 2-dimensional distribution in a flat-bottomed quadratic potential, as derived in Appendix A. The PMFs were defined to zero in the bulk region, and statistical errors of  $\leq 3 \text{ kJ mol}^{-1}$  were computed by bootstrapping complete histograms [181].

## 6.2 DRISM AND 3D-RISM CALCULATION DETAILS

Before the 3D-RISM calculations, DRISM calculations were carried out to obtain the bulk solvent susceptibility function  $\chi_{\gamma\alpha}(\mathbf{r})$ . We used the same solute parameters as in the MD simulations except for water and  $\text{O}_2$  solvent molecules. Water parameters were described by the TIP3P RISM model provided with the AMBER software [198], which, in contrast to the default TIP3P model, assumes LJ interactions for the hydrogen atoms. For  $\text{O}_2$ , we used the LJ parameters from the CHARMM22 force-field [117], and the partial charges were set to zero.

For DRISM calculations the `rism1d` program in the AmberTools 14 molecular modeling package [198, 203] was used, at a temperature of 300 K and a dielectric constant of  $\epsilon = 78.497$ . We used a grid of 32,768 points with a grid spacing of 0.025 Å. All solvent mixtures were defined using a water concentration of 55.5 M and a solute concentration of 0.5 M. We used the MDIIS accelerated numerical solver [213] with 20 vectors, a step size of 0.3 and a residual tolerance of  $1 \cdot 10^{-12}$ , and the KH closure relation. Table 6.1 summarizes the closure relations and tolerances used in DRISM calculations. Note that KH was the standard closure unless stated otherwise.

For 3D-RISM calculations, we used the crystal structures of UT-B and AQP1 (PDB IDs: 4EZC and 1J4N, respectively) [64, 214]. Protein atom interactions were described by the Amber ff99SB\*-ILDN force-field [109]. To test the influence of protein fluctuations on PMFs, we also took 100 UT-B structures and 100 AQP1 structures from an equilibration simulation, and 100 UT-B structures from an umbrella simulation in which urea is bound to UT-B.

We computed 3D-RISM solutions using the `rism3d.snglpnt` program of the AmberTools 14 molecular modeling package [198, 203], at a

temperature of 300 K. For UT-B, we used the trimer structure and a grid of  $240 \times 224 \times 180$  points in a box of size  $120 \times 112 \times 90 \text{ \AA}^3$ . For AQP<sub>1</sub>, we used the monomer structure, in a grid of  $160 \times 140 \times 192$  points and a box of  $80 \times 70 \times 96 \text{ \AA}^3$  (see Figure 6.1 B). In both systems, the distance of grid boundary to the protein was at least 15 Å. The MDIIS solution converged with a residual error between  $10^{-5}$  and  $2 \cdot 10^{-3}$  using 5 vectors, a step size of 0.7, and the KH closure. To test the influence of other closures, we also converged the 3D-RISM solutions with the chained Partial Series Expansion of order- $n$  (PSE- $n$ ): i. e. PSE-1/PSE-2, and PSE-1/PSE-2/PSE-3, as suggested by Giambaşu *et al.* [215]. Table 6.2 summarizes the closure relations and tolerances used to converge the 3D-RISM solutions for each solvent system.

### 6.2.1 PMF construction after 3D-RISM

The membrane normal  $z$  was taken as the reaction coordinate, where  $z = 0$  was defined either as the COM of the transmembrane region of UT-B or as the COM of the NPA motif of AQP<sub>1</sub>. The PMFs were calculated from 3D-RISM results using the 3-dimensional density function (Equation 5.34). The  $\alpha$  sites considered for each solute were: oxygen atom for water, nitrogen atom for NH<sub>3</sub>, carbon atom for urea and methanol, and oxygen atom for O<sub>2</sub>. (Note that since the density is normalized to unity in the bulk, the number of atoms in the solute molecule is irrelevant.)

In the bulk water and in the vestibule regions, the integration area  $A_g(z)$  from Equation 5.35 was taken as a square with an area of  $(1.6 \text{ nm})^2$ , which is smaller than the monomer area in both proteins.  $A_g$  was centered along the respective monomer center (see Figure 6.1 B/C). In the pore region  $|z| < 1 \text{ nm}$ ,  $A_g(z)$  was taken as a square centered along the channel center, with a side length  $2(r_c(z) + 3 \text{ \AA})$ , where  $r_c(z)$  is the channel radius. The channel centers and radii were computed with PROPORES [216]. Since 3D-RISM results were calculated with the protein in vacuum and not embedded in a lipid membrane, thereby generating unphysical solvent density in the lipid membrane region, the integration area  $A_g$  was forced to be large enough to account for the entropic effect from the tightening of the pore with respect to the bulk, and small enough to exclude the lipid membrane region. In addition, the integration was restricted to a smaller square in the pore region, thus excluding solvent droplets inside the protein matrix that are separated from the channel and do not contribute to permeation.

To average over multiple monomers or structures from simulation frames, we first averaged the density before computing the PMF from Equation 5.36, which yields

$$\Delta G_\alpha(z) = -k_B T \ln \frac{\langle g_\alpha^{uv}(z) \rangle_c}{g_\alpha^{uv}(z_{\text{bulk}})}, \quad (6.5)$$

where  $\langle \cdot \rangle_c$  represents the average over monomers and/or simulation frames. To facilitate the direct comparison between the PMFs obtained with 3D-RISM and the ones obtained with US MD simulations, we used the same trapezoidal correction described in Section 6.1.2, using a correction of  $4.21 \text{ kJ mol}^{-1}$  and  $3.47 \text{ kJ mol}^{-1}$  for UT-B and AQP<sub>1</sub>, respectively, given by the bulk area of  $A_g(z)$  and the monomer area. Thus, the 3D-RISM PMFs shown correspond to a channel density of one channel per cross section area of the monomer.

### 6.3 IC<sub>50</sub> ESTIMATE

Following the calculation in ref. [217], the IC<sub>50</sub> value of a solute bound to one of the channel binding sites is given by

$$\text{IC}_{50} = \left( A_{\text{mono}} \int_{\text{ch}} \exp \left[ -\frac{\Delta G(z)}{k_B T} \right] dz \right)^{-1}, \quad (6.6)$$

where the integral is taken across the channel. The symbol  $\Delta G(z)$  represents the PMF that is defined to zero in bulk water.

### 6.4 WATER OCCUPANCY INSIDE THE CHANNEL FROM A PMF

Since the calculated PMFs correspond to a defined reference area in the bulk, they can be translated into the equilibrium 1-dimensional solute density along the channel coordinate  $z$  at a given solute concentration in the bulk. Hence, the PMFs of water permeation can be translated into the equilibrium water occupancy of the channels. In bulk, the water 1-dimensional density

$$\left( \frac{\Delta N_w}{\Delta z} \right)_{\text{bulk}} \quad (6.7)$$

is given by i) the reference area of the PMF, i. e. the cross section area of one monomer  $A_{\text{mono}}$  ( $10.3 \text{ nm}^2$  and  $13.9 \text{ nm}^2$  for AQP<sub>1</sub> and UT-B, respectively), and ii) the number density of water ( $33.4 \text{ nm}^{-3}$ ). As a result, we obtained for Equation 6.7 the values  $344.3 \text{ nm}^{-1}$  and  $463.6 \text{ nm}^{-1}$  for AQP<sub>1</sub> and UT-B, respectively. Because the PMF was defined to be zero in the bulk, the 1-dimensional water density along the pore is

$$\left( \frac{\Delta N_w}{\Delta z} \right) (z) = \left( \frac{\Delta N_w}{\Delta z} \right)_{\text{bulk}} \exp \left[ -\frac{\Delta G(z)}{k_B T} \right]. \quad (6.8)$$

Integrating  $(\Delta N_w/\Delta z)(z)$  across the channel lumen yields the equilibrium water occupancy, that is the average number of water molecules inside the channel.

## 6.5 SYSTEM SET-UP FOR MD SIMULATIONS WITH FLUC-BPE

6.5.1 *Amber ff99SB\*-ILDN force-field*

Fluc-Bpe initial coordinates were taken from Stockbridge *et al.* (PDB ID: 5NKQ) [85]. Lipid bilayer initial coordinates were taken from a 1 ns equilibrium simulation of a 328 POPC lipid hydrated bilayer. The protein was embedded in the bilayer with the `g_membed` software [45]. The simulation box was filled by explicit TIP3P [200] water molecules and neutralized with counter ions. Four different simulation systems were set up either of the WT or of the F82I/F85I mutant, each including or excluding the pore-blocking L2 monobodies. The final simulation system with pore-blocking L2 monobodies consisted in one homodimeric Fluc-Bpe, two pore-blocking L2 monobodies, 188 POPC lipids, 17685 water molecules, 6 Na<sup>+</sup> ions, and 4 F<sup>-</sup> ions (see Figure 6.2A). In the same fashion, the final simulation system without L2 monobodies consisted in one homodimeric Fluc-Bpe, 204 POPC lipid molecules, 18052 water molecules, 1 Na<sup>+</sup> ion, and 7 F<sup>-</sup> ions (Figure 6.2C). The ion parameters were taken from Joung & Cheatham III (2008) [218]. The F82I/F85I mutant was generated using PyMOL [219].

Protein atom interactions were described by the Amber ff99SB\*-ILDN force-field [109], and lipid parameters were taken from Berger *et al.*, and Cordoní *et al.* [201, 202]. The simulations were performed using the Gromacs simulation software (version 5.1.4) [152]. Electrostatic interactions were calculated at each step with PME [163]. Short-range repulsive and attractive dispersion interactions were described by a LJ potential, with a cut-off at 1 nm. The geometry of water molecules was constrained with the SETTLE algorithm [155], and all other bond lengths were constrained with LINCS [157]. Hydrogen atoms of the protein were constructed as virtual sites, allowing a 4 fs time step [158].

The temperature in the simulation was kept constant at 300 K using velocity rescaling ( $\tau = 2.5$  ps) [164], and the pressure was kept at 1 bar with a semiisotropic Berendsen barostat ( $\tau = 2$  ps) [165]. The systems were equilibrated for 20 ns with position restraints on the protein heavy atoms and on the molecules placed at the electron density sites (F<sup>+</sup> or water molecules at F82, and F85 sites (see Figure 6.2B), with a force constant of  $1000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$ . Afterwards, the systems were equilibrated for 400 ns with position restraints ( $k = 1000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$ ) on the protein C <sub>$\alpha$</sub>  atoms and on the molecules placed at F82, and F85 sites before production.

To test the stability of the F<sup>-</sup> and water molecules assigned to the densities at F82 and F85 sites, multiple unconstrained simulations were started from random time points taken from the 400 ns equili-

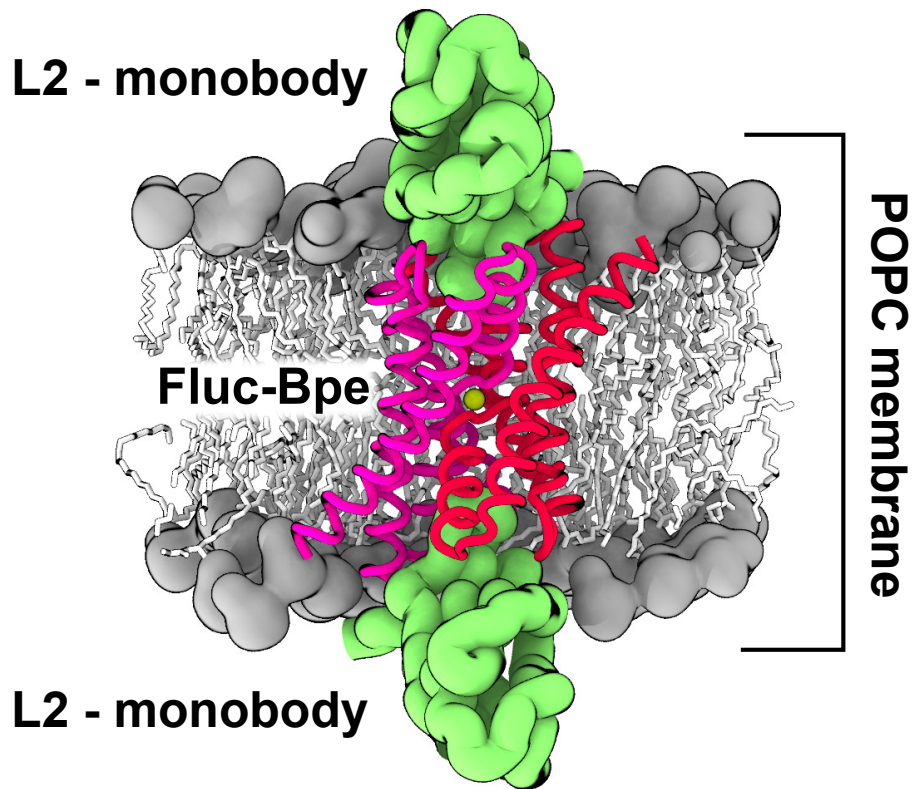


Figure 6.2: **Simulation box with Fluc-Bpe.** The membrane lipid heads are represented in grey volume, and the lipid tails as white sticks. Fluc-Bpe is depicted as red/magenta ribbons, and the L2 monobodies are coloured in green. The structural  $\text{Na}^+$  ion is depicted in yellow.

bration simulations. Here, four different system types were defined for unconstrained simulations:

- Four  $\text{F}^-$  at F82 and F85 sites / without pore-blocking monobodies.
- $\text{F}^-$  at F82 and F85 sites / with pore-blocking monobodies.
- Four water molecules at F82 and F85 sites / without pore-blocking monobodies.
- Four water molecules at F82 and F85 sites / with pore-blocking monobodies.

For systems with or without pore-blocking L2 monobodies, we performed 10 replica simulations of 80 ns each or 20 replica simulations of 40 ns each, respectively. All the unconstrained simulations were carried out without position restraints, and using the parameters described above.

Likewise, we tested the protein structural stability under fluoride binding conditions. Production simulations were started from ran-



dom time points taken from the 400 ns equilibration simulation without pore-blocking monobodies. The fluoride ions were restrained with a force constant  $k = 1000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$  at F82 and F85 sites. 20 replica simulations of 80 ns each were carried out, using the parameters described above.

### 6.5.2 US simulation details

The reaction coordinate was defined as the  $z$ -coordinate (membrane normal), measured as the distance between the COM of the solute molecule to COM of the transmembrane backbone atoms of the pore being permeated. Starting structures for US were taken from the last 10 ns of a 90 ns simulation in which fluoride ions were restrained at F82 and F85. We placed 3 solute molecules per simulation, from which 3 umbrella windows were collected per simulation. Each solute was placed at a distance of 3 nm from each other in the  $z$ -direction, and restrained by a harmonic umbrella potential with a force constant of  $5000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$ . The solute molecules were restrained via a cylindrical flat-bottomed potential (see Equation 6.1) to keep them within a certain area outside the channel (bulk area), with a force constant  $k_c = 5000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$ , and a cylinder radius  $r_c = 0.8 \text{ nm}$ . All the US simulations were performed using the Gromacs simulation software (version 5.1) [152]. Adjacent umbrella windows were separated by 0.05 nm, needing 60 simulation systems to cover the reaction coordinate. US simulations were conducted for 75 ns each, in which the temperature was controlled using a stochastic dynamics integration scheme ( $\tau = 0.5 \text{ ps}$ ) [211], and the pressure was controlled at 1 bar with a semiisotropic Parrinello-Rahman barostat ( $\tau = 5 \text{ ps}$ ) [166, 167]. All other simulation parameters were identical to the equilibrium simulations.

### 6.5.3 PMF construction after US

The first 25 ns of each trajectory was removed for equilibration in simulations with water and chloride, and the first 40 ns in simulations with fluoride. A total of 540 histograms were collected from  $13.5 \mu\text{s}$  of simulation with the Fluc-Bpe system. We used a periodic implementation of WHAM to calculate the PMFs [179, 181], which were defined to zero in the bulk region. The complete histograms were bootstrapped yielding statistical errors of  $\leq 8 \text{ kJ mol}^{-1}$  [181].

### 6.5.4 CHARMM36 force-field

To rule out a possible effect due to the applied protein force-field in the qualitative results, a different set of simulation systems were set-up with the CHARMM force-field family [122, 220]. The system

set-up was identical to the set-up for the Amber force-field (see Section 6.5.1), with the only difference that Fluc-Bpe was solvated in a membrane of 1,2-dipalmitoyl-*sn*-phosphatidylcholine (DPPC) instead of POPC. The final CHARMM36 simulation system consisted in one homodimeric WT Fluc-Bpe, 228 DPPC lipid molecules, 13273 TIP3P water molecules, 1 Na<sup>+</sup> ion, 3 Cl<sup>-</sup> ions, and 4 F<sup>-</sup> ions. The F<sup>-</sup> ion parameters were again taken from ref. [218], and the Na<sup>+</sup> and Cl<sup>-</sup> parameters were taken from the CHARMM36 default parameters.

Protein and lipid atom interactions were described by the CHARMM36 force-field [122, 220]. The simulations were carried out with the Gromacs simulation software (version 5.1.4) [152]. Electrostatic interactions were calculated at every step with the PME method [163], and using a cut-off at 1.2 nm for LJ interactions. The geometry of water molecules was constrained with the SETTLE algorithm [155], and all hydrogen bonds were constrained with LINCS [157]. A 2 fs timestep was used. The simulation temperature was kept constant at 323 K using velocity rescaling ( $\tau = 2.5$  ps) [164], and the pressure was kept at 1 bar with a semiisotropic Berendsen barostat ( $\tau = 2$  ps) [165]. The system was equilibrated for 20 ns with position restraints on the protein heavy atoms and on the F<sup>-</sup> ions placed at F82 and F85 with a force constant of  $k = 1000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$ . After that, the system was equilibrated for 5 ns with position restraints ( $k = 1000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$ ) on the protein C $_{\alpha}$  atoms and on the F<sup>-</sup> ions placed at F82 and F85 before production.

Finally, by assigning random starting velocities, we generated 20 independent replicas from the last snapshot taken from the 5 ns equilibration trajectory, and we ran 10 unconstrained replicas for 40 ns, and another 10 replicas, with position restraints on F<sup>-</sup> at F82 and F85, for 80 ns.

### 6.5.5 Polarizable DRUDE force-field

To investigate the effect of polarizability on the qualitative results, a simulation system was set-up using the polarizable Drude force-field (see Section 5.1.3.2) [31, 147, 148, 221]. Initial coordinates for the lipidic bilayer were taken from a 3 ns equilibrium simulation of a 256 DPPC lipid hydrated bilayer, and the protein was embedded in the membrane with the `g_membed` software [45]. The simulation box was filled by explicit SWM4-NDP [222] water and neutralized with counter ions. The final simulation system consisted of one homodimeric WT Fluc-Bpe, 228 DPPC lipids, 10181 water molecules, 1 Na<sup>+</sup>, and 7 F<sup>-</sup> ions. The ion parameters were taken from Yu *et al.* (2010) [223].

Protein atom interactions were described by the polarizable CHARMM-Drude force-field [31, 147, 148, 221], and lipid parameters were taken from ref. [224]. The Gromacs DPPC topology file was generated from

original CHARMM-Drude topology files via TopoGromacs [225]. The simulations were performed using the Gromacs simulation software (version Drude) [226, 227]. Electrostatic interactions were calculated at every step with the PME method [163], and using a cut-off at 1.2 nm for LJ interactions. The geometry of water molecules was constrained with the SETTLE algorithm [155], and all other bond lengths were left unconstrained.

The system was carefully equilibrated with multi-step protocol to avoid the polarization catastrophe. First, we equilibrated the simulation system using the SCF approach, as implemented by Van Maaren & Van Der Spoel [226]. The simulation temperature was kept constant at 323 K using velocity rescaling ( $\tau = 0.1$  ps) [164], and the pressure was kept at 1 bar with a semiisotropic Berendsen barostat ( $\tau = 1$  ps) [165]. Using a 0.2 fs time step, the system was equilibrated for 37 ps with position restraints on the protein heavy atoms ( $k = 1000$  kJ mol<sup>-1</sup> nm<sup>-2</sup>). After that, the system was equilibrated for 800 ps with position restraints on the protein C<sub>α</sub> atoms and on the F<sup>-</sup> placed at F82 and F85. In this last equilibration, the simulation temperature was kept constant at 323 K for the non-drude atoms, and at 1 K for the drude particles, using the Nose-Hoover thermostat ( $\tau = 0.1$  ps, and  $\tau = 0.005$  ps, respectively) [228, 229]. No pressure-coupling was used, and a 1 fs time step was used with the extended Lagrangian dynamics implemented in Gromacs by Lemkul *et al.* [227].

Finally, by generating random starting velocities, we generated 10 different replicas from the last snapshot taken from the 800 ps equilibration trajectory, and ran each of them for 10 ns.

## 6.6 FREE-ENERGY CALCULATIONS OF ABSOLUTE AND RELATIVE BINDING IN FLUC-BPE

We carried out two different sets of free energy calculations: *relative* binding free energies and *absolute* binding free energies. The *relative* binding free energy difference between two putative molecular species determines which species is more likely to bind a certain binding site. On the other hand, the *absolute* binding free energy determines whether a certain molecular species can actually bind to a certain binding site.

**RELATIVE BINDING FREE ENERGIES** To compute *relative* binding free energies, a water molecule was alchemically transformed into a F<sup>-</sup> ion in bulk solvent, whilst simultaneously transforming a F<sup>-</sup> into a water molecule at the F82 or F85 binding site. According to the thermodynamic cycle depicted in Figure 6.3, the associated free energy difference equals the free energy difference  $\Delta\Delta G_{\text{bind}}$  of binding an ion versus a water molecule into the binding site. A negative  $\Delta\Delta G_{\text{bind}}$  implies that a water molecule is more likely to translocate

from solvent to the binding site as compared to the ion. More quantitatively, in equilibrium, the probabilities of finding a water molecule  $P_{\text{bind}}^{\text{wat}}$  or of finding an ion  $P_{\text{bind}}^{\text{ion}}$  at the binding site are related via

$$\frac{P_{\text{bind}}^{\text{wat}}}{P_{\text{bind}}^{\text{ion}}} = \frac{c_{\text{wat}}}{c_{\text{ion}}} \exp\left(\frac{-\Delta\Delta G_{\text{bind}}}{k_{\text{B}}T}\right), \quad (6.9)$$

where  $c_{\text{wat}}$  and  $c_{\text{ion}}$  represent the water and ion concentrations in the solvent. Hence, the  $\Delta\Delta G_{\text{bind}}$  is in principle the key quantity to detect the correct assignment of the ambiguous electron densities. However,  $\Delta\Delta G_{\text{bind}}$  is associated with an increased uncertainty, mainly because force-fields might not represent ion-protein interactions as accurately as water-protein interactions [230–234].

**ABSOLUTE BINDING FREE ENERGIES** We also calculated the *absolute* binding free energies  $\Delta G_{\text{bind}}$  for water molecules at F82 and F85 sites, i. e. the free energy for translocating a water molecule from solvent into the binding site. Since the probability of finding a water molecule at the binding site is proportional to  $\exp(-\Delta G_{\text{bind}}/k_{\text{B}}T)$ , the assignment of a water molecule to an ambiguous electron density is only plausible if  $\Delta G_{\text{bind}}$  is negative or at least close to zero.

All free energies were calculated via discrete TI along an alchemical reaction coordinate  $\lambda$ , where  $\lambda = 0$  and  $\lambda = 1$  represent the initial and the final state, respectively. During TI calculations, a time step of 2 fs was used, and the temperature was kept constant at 300 K using a stochastic dynamics integration scheme ( $\tau = 0.1$  ps) [211]. Starting configurations were taken from the final snapshot of the 20 ns equilibrium simulation. All the other simulation parameters were identical to the equilibrium simulations (see Section 6.5.1).

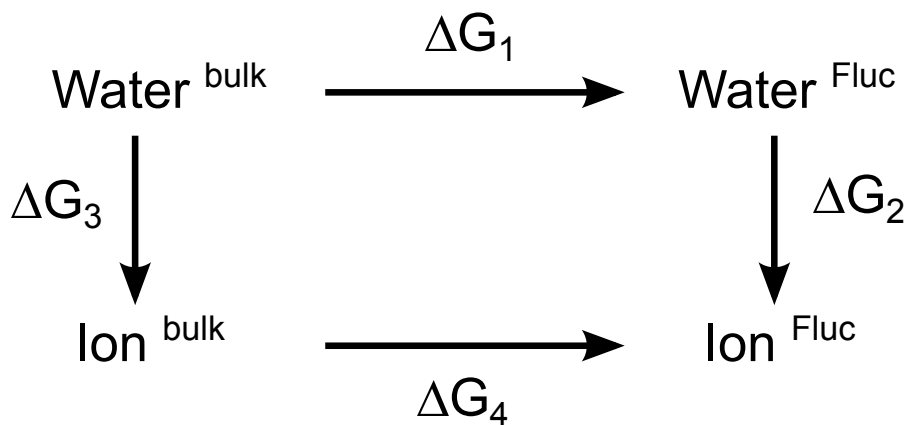


Figure 6.3: **Alchemical transformation thermodynamic cycle for computing the relative binding free energy.** From the thermodynamic cycle, the binding free energy difference  $\Delta\Delta G_{\text{bind}} = \Delta G_1 - \Delta G_4$  between the ion-bound Fluc channel and the water-bound Fluc channel can be calculated via  $\Delta G_3 - \Delta G_2$ .

### 6.6.1 Relative binding free-energy computational details

In the relative binding free energy calculations,  $\lambda = 0$  described the state with the ion in the protein ( $F^-$  at F82 or at F85, or  $Na^+$  TM3) while having a water molecule in bulk solvent.  $\lambda = 1$  described the state with the ion ( $F^-$  or  $Na^+$ ) in bulk solvent while having a water molecule at the F82, F85, or TM3. The free energy calculations with  $F^-$  were done in the presence or absence of pore-blocking L2 monobodies, whereas the free energy calculations with  $Na^+$  were done only in presence of monobodies. To keep the protein structure close to the crystal structure, positional restraints ( $k = 1000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$ ) were applied on the protein  $C_\alpha$  atoms and on the ion and water molecules that were alchemically transformed. The molecule that was restrained in bulk solvent was restrained at a position as far as possible from the protein and membrane. To avert numerical instabilities close to  $\lambda = 0$  and  $\lambda = 1$ , we used soft-core potentials for both electrostatics and LJ interactions with  $\alpha = 0.5$ ,  $\sigma = 0.3$  and a soft-core power of 1. The transformation from  $\lambda = 0$  to  $\lambda = 1$  was done in 41  $\lambda$ -steps, simulating each  $\lambda$ -step for 10 ns. Derivatives of the Hamiltonian  $H$  with respect to  $\lambda$  were recorded at every step. Free energy differences were calculated from

$$\Delta\Delta G_{\text{bind}} = \int_0^1 d\lambda \left\langle \frac{\delta H}{\delta \lambda} \right\rangle \quad (6.10)$$

where  $\langle \cdot \rangle$  is the average over the trajectory. The first 50 ps of each trajectory were removed for equilibration.

### 6.6.2 Absolute binding free energy computational details

We calculated the absolute binding free energies  $\Delta G_{\text{bind}}$  of water via TI as depicted in Figure 6.4. We calculated  $\Delta G_{\text{bind}}$  from the difference in free energy between i) turning off the interactions of a water molecule in a pure water box ( $\Delta G_2$  in Figure 6.4), and ii) turning off the interactions of a water molecule in the WT Fluc-Bpe at F82 and F85, or in the mutant Fluc-Bpe at F82I and F85I sites ( $\Delta G_1$  in Figure 6.4). Since there are two F82 sites and two F85 sites in Fluc, the absolute  $\Delta G_{\text{bind}}$  of water to each of the two binding sites was calculated.

As indicated by Deng & Roux [189], when calculating absolute binding free energies, it is desirable to restrain the ligand at the binding site to make the sampling more efficient. Thus, to keep the water molecules at the positions indicated by the electron densities in the crystal structure, we applied position restraining potentials to the oxygen atom of water molecules at Fluc-Bpe binding sites. Likewise, we also applied position restraints on the water molecule in the pure water box. Following the scheme shown in Figure 6.4, the calculation of  $\Delta G_1$  was separated in five sub-steps:

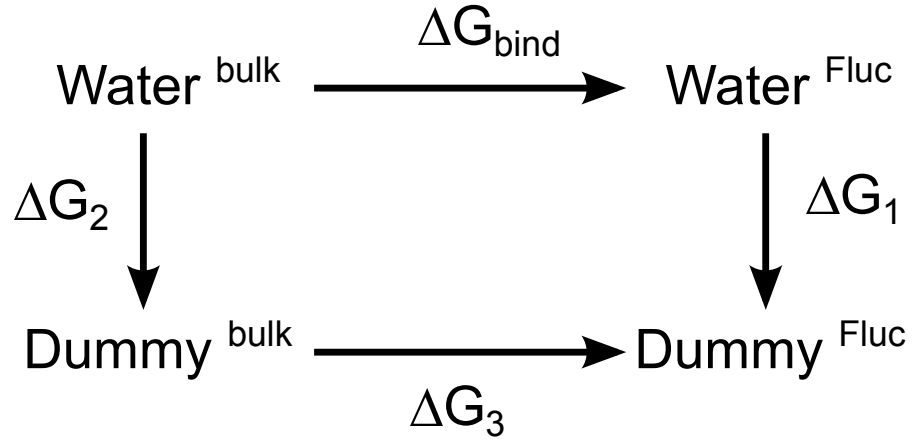


Figure 6.4: **Absolute binding free energy thermodynamic cycle.** From the thermodynamic cycle, the absolute binding free energy is calculated via  $\Delta G_{\text{bind}} = \Delta G_2 - \Delta G_1$ , where  $\Delta G_3 \equiv 0$ .

- $\Delta G_1^{\text{posre}}$ , turning on the position restraints with a force constant  $K_x = 200 \text{ kJ mol}^{-1} \text{ nm}^{-2}$ .
- $\Delta G_1^{\text{Coulomb}}$ , turning off electrostatic interactions.
- $\Delta G_1^{\text{LJ}}$ , turning off LJ interactions.
- Translating the non-interacting solute from the binding site into a pure-solvent box.
- Release of the positional restraints in pure solvent,  $\Delta G_{r,x}^\circ$

To obtain  $\Delta G_1^{\text{posre}}$ , 11  $\lambda$ -steps were simulated, 2 ns each, for all systems.

$\Delta G_1^{\text{Coulomb}}$  was computed from 11  $\lambda$ -steps simulated for 2 ns each, and  $\Delta G_1^{\text{LJ}}$  was computed in 41  $\lambda$ -steps simulated for 2 ns each. To account for atomic overlaps occurring close to  $\lambda = 0$  and  $\lambda = 1$ , soft-core potentials were used for LJ interactions with  $\alpha = 0.5$ ,  $\sigma = 0.3$  and a soft-core power of 1. As before, the first 50 ps of each trajectory were removed for equilibration, and the free energy differences were calculated according to Equation 5.29. The calculation of  $\Delta G_2$  in Figure 6.4 was separated into the electrostatic part and LJ interactions part, and the free energy  $\Delta G_{r,x}^\circ$  due to applying a position restraining potential on a non-interacting water particle was calculated analytically as in ref. [235]. To apply a translational restraining potential  $U_{r,x}$  with a force constant  $K_x$  and the form

$$U_{r,x} = \frac{K_x(\mathbf{x} - \mathbf{x}_0)^2}{2} \quad (6.11)$$

confines the molecule to a small effective volume [235]:

$$V_{\text{eff}} = \int_V \exp \left[ \frac{-1}{k_B T} \frac{K_x(\mathbf{x} - \mathbf{x}_0)^2}{2} \right] d\mathbf{x}, \quad (6.12)$$

that can be analytically evaluated as

$$V_{\text{eff}} = \sqrt{\left(2\pi \frac{k_B T}{K_x}\right)^3}. \quad (6.13)$$

Thus, the corresponding free energy can be calculated via

$$\Delta G_{r,x}^\circ = -k_B T \ln \left( \frac{V^\circ}{V_{\text{eff}}} \right), \quad (6.14)$$

where  $V^\circ = 0.02992 \text{ nm}^3$  is the molecular volume of water [236]. We used  $K_x = 200 \text{ kJ mol}^{-1} \text{ nm}^{-2}$ , which yields  $\Delta G_{r,x}^\circ = 0.774225 \text{ kJ mol}^{-1}$ .

## 6.7 MAXIMUM-LIKELIHOOD ESTIMATES

To calculate the lifetime  $\tau$  of the fluoride or water molecules at F82 and F85 binding sites in Fluc, we used a maximum-likelihood estimate as described recently in ref. [237]. In summary, given  $M$  simulations of simulation time  $T_s$ , during which  $m$   $F^-$  or water molecules leave the binding site at times  $t_i$  ( $i = 1, \dots, m$ ;  $m \leq M$ ), then the maximum-likelihood estimate  $\tau_{\text{ml}}$  for the lifetime of  $F^-$  or water at the binding site is given via

$$\tau_{\text{ml}} = \frac{(M - m)T_s + \sum_{i=1}^m t_i}{m}. \quad (6.15)$$

## 6.8 QUANTUM CHEMICAL CALCULATIONS

The relaxed Potential Energy Surface (PES) calculations of  $F^-$  and benzene were computed via the DFT-B3LYP functional [238] or the Møller-Plesset 2<sup>nd</sup>-order (MP2) perturbation theory [239] as implemented in the ORCA 3 program system [240]. The PESs calculated with the DFT-B3LYP functional were obtained using either the ma-def2-SVP or the ma-def2-TZVP basis sets [241, 242]. Additionally, MP2 calculations were performed with the aug-cc-pVDZ basis set [243].

DRISM

Solvent	Closure	Tolerance
water 55.5M	KH, PSE-2, PSE-3, HNC	$10^{-12}$
water 55.5M + NH <sub>3</sub> 0.5M	KH	$10^{-12}$
water 55.5M + urea 0.5M	KH, PSE-2, PSE-3, HNC	$10^{-12}$ , $10^{-3}$ , $10^{-3}$ , $10^{-3}$
water 55.5M + O <sub>2</sub> 0.5M	KH	$10^{-12}$
water 55.5M + methanol 0.5M	KH	$10^{-12}$

Table 6.1: Different closure relations and residual errors used to converge DRISM calculations with each solvent system.

3D-RISM

Solvent	Closure	Tolerance
water 55.5M	KH, PSE-2, PSE-1/PSE-2/PSE-3	$10^{-5}$
water 55.5M + NH <sub>3</sub> 0.5M	KH, PSE-1/PSE-2, PSE-1/PSE-2/PSE-3	$10^{-5}$ , $10^{-5}$ / $6 \cdot 10^{-3}$ / $10^{-2}$
water 55.5M + urea 0.5M	KH, PSE-1/PSE-2	$5 \cdot 10^{-3}$ , $5 \cdot 10^{-3}$ / $5 \cdot 10^{-2}$
water 55.5M + O <sub>2</sub> 0.5M	KH, PSE-1/PSE-2, PSE-1/PSE-2/PSE-3	$10^{-5}$ , $10^{-5}$ / $10^{-5}$ / $10^{-4}$
water 55.5M + methanol 0.5M	KH, PSE-1/PSE-2	$5 \cdot 10^{-3}$ , $3 \cdot 10^{-3}$ / $10^{-2}$

Table 6.2: Different closure relations and residual errors used to converge 3D-RISM calculations with each solvent system.



Part III

RESULTS



---

## A COMPARISON BETWEEN MD AND 3D-RISM

---

**P**ROTEIN CHANNELS assist the selective permeation of molecules across membranes. A common method to study such process is the computation of the Potential of Mean Force (PMF) (see Section 5.2) of solute permeation across the channel pore. PMFs are typically calculated from Molecular Dynamics (MD) simulations, often using enhanced sampling methods such as Umbrella Sampling (US). In this chapter we present the results of PMF calculations for water, ammonia, urea, molecular oxygen, and methanol across Urea Transporter B (UT-B) and Aquaporin 1 (AQP1), using MD simulations with US, as well as with the more computationally efficient and accurate 3-Dimensional Reference Interaction-Site Model (3D-RISM). While the calculation of PMFs using MD simulations is a well established procedure, the suitability of 3D-RISM-based PMF calculations has remained controversial. Therefore, we aim to compare the PMFs calculated from 3D-RISM results to the PMFs calculated from MD simulations, and to find out whether 3D-RISM may be considered as a computationally efficient replacement of MD simulations for solute permeation studies.

### 7.1 PMF CALCULATIONS FROM UMBRELLA SAMPLING

#### 7.1.1 UT-B

Figure 7.1 A shows the PMFs  $\Delta G(z)$  for the permeation of water,  $\text{NH}_3$ , urea,  $\text{O}_2$ , and methanol across UT-B, which were computed as described in Section 6.1. These PMFs are consistent with previous calculations [63, 64, 244]. The free energy barriers  $\Delta G^\ddagger \leq 15 \text{ kJ mol}^{-1}$  suggest that UT-B facilitates the flux of small neutral solutes in a non-selective fashion. The highest free energy barriers are observed in the  $S_m$  region, at  $z \approx 0$ , where the hydrophobicity of the pore hinders the permeation of the polar solutes: water,  $\text{NH}_3$ , and urea.  $\text{O}_2$  and methanol, being more hydrophobic, exhibit lower free energy barriers  $\Delta G^\ddagger$  in the  $S_m$  region.

Next to the  $S_m$  region, the PMF for urea (Figure 7.1 A, green line) reveals several energy minima indicating the urea binding sites. Our PMF for urea permeation is similar to the PMF presented by Levin *et al.*,

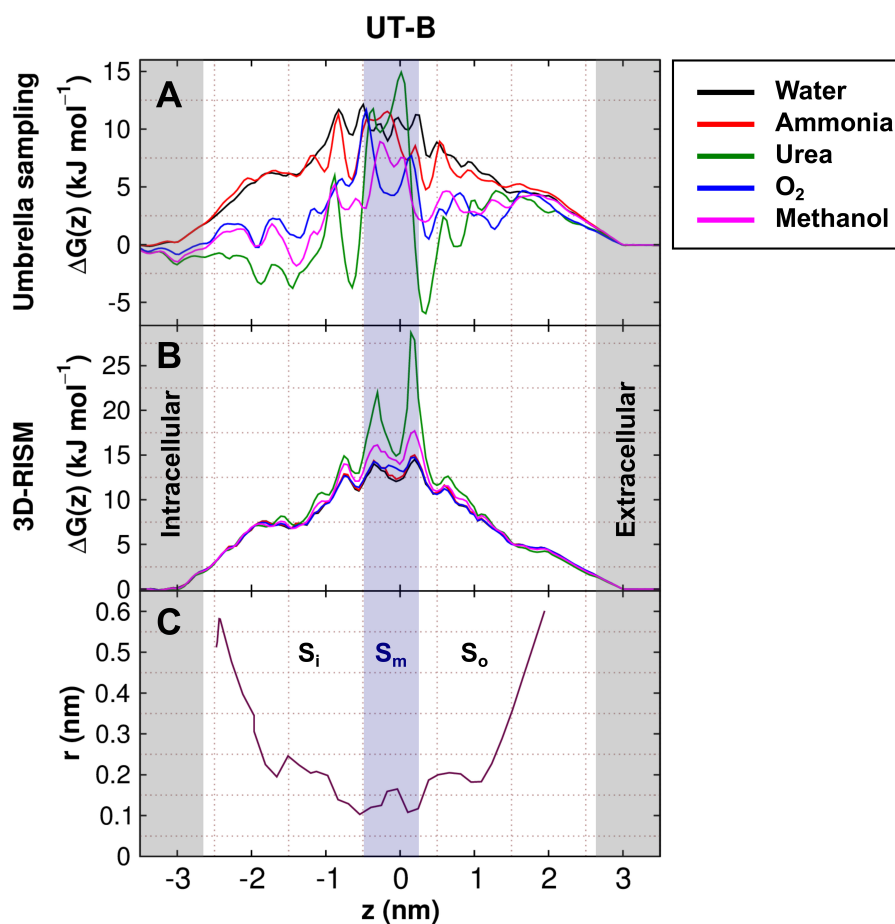


Figure 7.1: PMFs from Umbrella Sampling and 3D-RISM for UT-B. PMFs  $\Delta G(z)$  for solute permeation along the channel coordinate  $z$  from US simulations for UT-B, calculated with (A) US, and (B) 3D-RISM. Different solutes are represented by different colors (see legend). (C) Radius of UT-B computed with PROPORES [216].  $z = 0$  corresponds to the COM of the transmembrane residues in UT-B.

in which the binding sites for crystallized selenourea in  $S_i$  ( $z = -1$  nm) and  $S_o$  ( $z = 0.4$  nm and  $z = 0.6$  nm) regions in the UT-B crystal are indicated (see also Figure 2.2) [64]. Surprisingly, no selenourea was bound to the crystal at the energy minimum at  $z = -0.6$  nm, probably indicating that the packing of selenourea to the protein is different compared to that of urea.

To compare the urea concentration at the binding site with bulk water, we computed the  $IC_{50}$  concentration for urea binding as explained in Section 6.3. The resulting  $IC_{50} = 20$  mM suggests that urea occupies a substantial fraction of UT-B channels at physiological concentrations [245].

The narrowing of the pore reduces the space sampled by the solute, which leads to an entropic barrier. Hence, the free energy barriers  $\Delta G^\ddagger \sim 12$  kJ mol<sup>-1</sup> in the PMFs for water and NH<sub>3</sub> reveal that the permeation of such solutes is a mainly entropic process as the molecules enter the narrow lumen of the pore (see Figure 7.1 A, black and red curves, respectively, and Figure 7.1 C). These rather low free energy barriers suggest that UT-B is an efficient water and NH<sub>3</sub> channel, which is in agreement with previous studies [62, 63, 244]. Similarly, the energy barriers observed in the PMFs for the permeation of O<sub>2</sub> and methanol are also low, suggesting that they might as well permeate UT-B. It is important to stress that each PMF shows a different profile, as a result of the different characteristics—shape and hydrophobicity—of each solute.

### 7.1.2 AQP1

Figure 7.2 A presents the PMFs  $\Delta G(z)$  for the permeation of water, NH<sub>3</sub>, urea, and O<sub>2</sub> across AQP1, which were taken from ref. [207] (see Hub & de Groot (2008) and the discussion therein) with the sole purpose of comparing them with the PMFs  $\Delta G(z)$  computed with 3D-RISM.

Briefly, the PMFs in Figure 7.2 A exhibit the highest energy barriers at the aromatic/arginine (ar/R) constriction region, indicating the selectivity filter of the channel. Small polar molecules permeate across this selectivity filter, whilst hydrophobic and large ones are excluded. Since the ar/R region is the narrowest part of the pore (see Figure 7.2 C), large solutes such as urea remain excluded due to steric hindrance. Small hydrophobic solutes such as O<sub>2</sub> will not permeate unless they displace the water molecules that establish strong hydrogen bonds with the conserved arginine residue [246]. Overall, these PMFs reflect the steric, hydrophobic and polar interactions by which the permeation of different solutes is modulated.

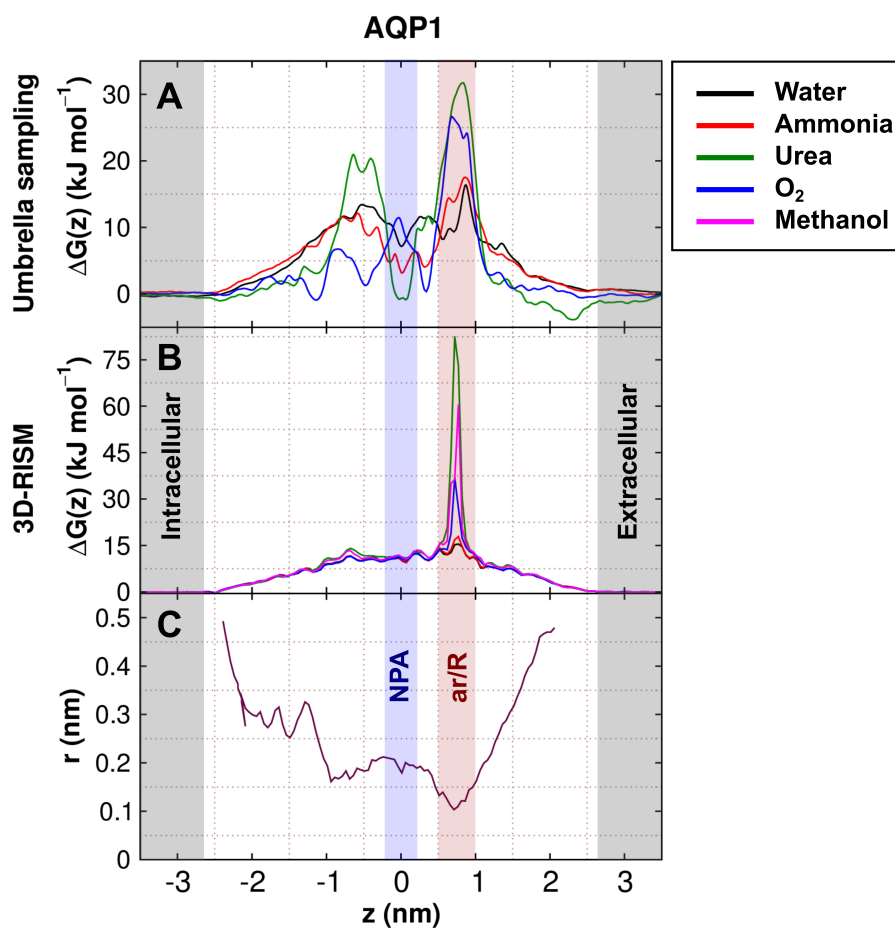


Figure 7.2: PMFs from Umbrella Sampling and 3D-RISM for AQP1. PMFs  $\Delta G(z)$  for solute permeation along the channel coordinate  $z$  from US simulations for AQP1, calculated with (A) US, and (B) 3D-RISM. Different solutes are represented by different colors (see legend). (C) Radius of AQP1 computed with PROPORES [216].  $z = 0$  corresponds to the COM of the NPA motif in AQP1.

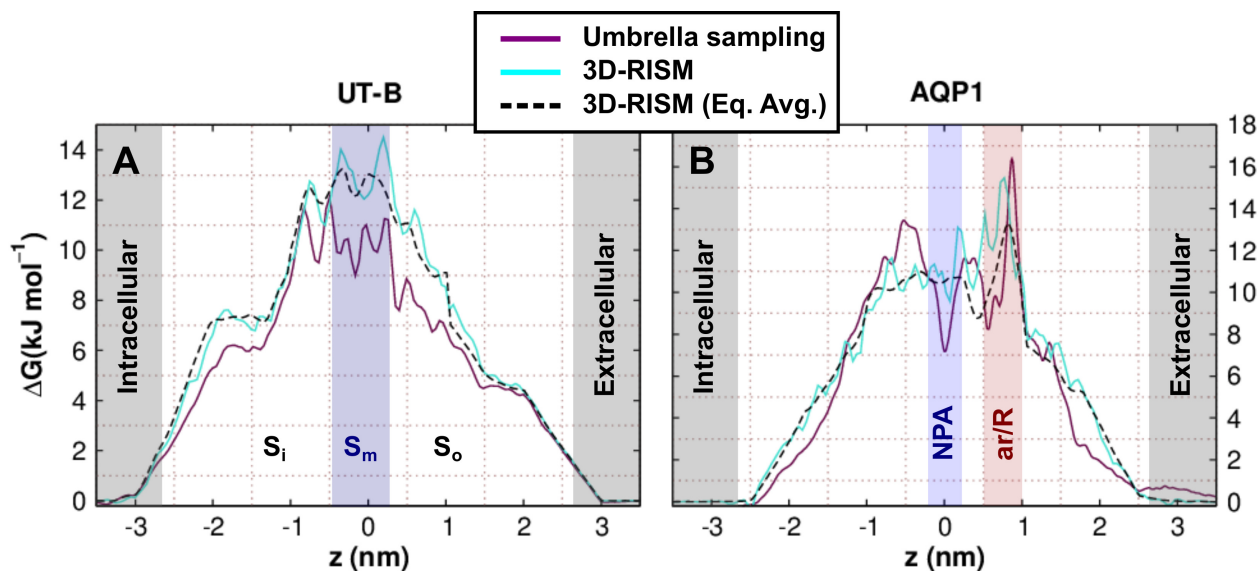


Figure 7.3: PMFs for water permeation across UT-B and AQP1. Water PMFs  $\Delta G(z)$  calculated with US simulations (brown), 3D-RISM with the crystal structure (turquoise) and 3D-RISM after averaging over 100 different structures taken from equilibrium MD simulations (black-dashed) for (A) UT-B and (B) AQP1. The  $S_i$ ,  $S_m$ , and  $S_o$  regions are highlighted for UT-B, as well as the NPA motif and the ar/R region for AQP1.

## 7.2 PMF CALCULATIONS FROM 3D-RISM

Here we compare the PMFs for the permeation of water,  $\text{NH}_3$ , urea,  $\text{O}_2$ , and methanol across UT-B and AQP1 calculated with US, with the PMFs computed with 3D-RISM (see Section 6.2). To allow quantitative comparison with the PMFs computed with US, the PMFs computed with 3D-RISM were normalized with respect to the monomer cross section area using the trapezoidal correction between the entrance and the exit regions of the pore (see Section 6.1.2) that was also applied to the PMFs from US simulations.

### 7.2.1 3D-RISM PMFs for water permeation are consistent with US PMFs

Figure 7.3 shows the PMFs for water permeation across UT-B and AQP1 computed with 3D-RISM and US simulations. For UT-B, 3D-RISM results in a PMF that exhibits higher free energy barriers as compared with US, which can be translated into a lower water density inside the pore. Nevertheless, differences of  $\lesssim 4 \text{ kJ mol}^{-1}$  between the two curves indicate a reasonable agreement between the two methods. To rule out that 3D-RISM results may be affected by the structural conformation of the crystal, we calculated, using 3D-RISM, the average PMFs over 100 structures taken from equilibrium MD simulations (black dashed curve in Figure 7.3). Once these structural fluctuations are consid-

	3D-RISM (cryst.)	3D-RISM (av.)	MD
UT-B	3.1	3.2	7.8
AQP1	8.3	10.4	10.4 (7.7)

Table 7.1: Average water occupancy in the narrowest region of the channel lumen of UT-B ( $-0.8 \text{ nm} < z < 0.3 \text{ nm}$ ) and AQP1 ( $-1 \text{ nm} < z < 1 \text{ nm}$ ) as computed from the PMFs. 3D-RISM results computed using only the crystal structure or from an average over 100 MD simulations frames. MD result in brackets was computed by counting water molecules during an equilibrium simulation, taken from ref. [207].

ered, the PMFs calculated with 3D-RISM yield differences  $\lesssim 2 \text{ kJ mol}^{-1}$  as compared with US, suggesting that the structural fluctuations of the solute (protein) may influence the 3D-RISM calculations of solvent density distributions.

To provide an additional quantitative comparison between the PMFs computed with US simulations and 3D-RISM, we calculated the average water occupancy in the narrowest part of each channel from the PMFs, as explained in Section 6.4. Table 7.1 presents the results obtained with this approach. The average water occupancies  $\sim 8$  suggest that a file of water molecules fills the channel pores in equilibrium MD simulations, which is in agreement with previous MD simulations [247–249]. Whereas the results for AQP1 are quantitatively consistent between 3D-RISM and US, the average water occupancies in UT-B, as calculated from the PMF obtained with 3D-RISM, are lower as compared with US. However, given that the 3D-RISM calculations are orders of magnitude faster than US simulations, the general correspondence is noteworthy.

### 7.2.2 *The solute size dominates the 3D-RISM PMFs*

In contrast, the PMFs for the permeation of non-water solutes, calculated with US and 3D-RISM, are not as consistent as the PMFs for water permeation (see Figure 7.1 B, and Figure 7.2 B). The PMFs computed from US simulations of different solutes greatly differ, reflecting their different chemical and steric features. However, the PMFs computed from 3D-RISM barely show any difference for different solutes. The stark resemblance between the PMFs for water,  $\text{NH}_3$ , urea,  $\text{O}_2$ , and methanol in each channel is only broken at the  $S_m$  region in UT-B and the ar/R constriction site in AQP1. In particular, the polarity of the solutes does not seem to be reflected on the PMFs. The PMFs yielded from 3D-RISM exhibit higher energy barriers the larger the solute, suggesting that the energy profiles are intensely dominated by the solute size. As a result, the urea binding sites in UT-B reported in this (Figure 7.1 A) and other computational studies [64] are not revealed by



in the PMF calculated from 3D-RISM (Figure 7.1 B, and Figure 7.2 B). The PMFs computed with 3D-RISM for urea, O<sub>2</sub>, and methanol show pronounced energy barriers at the ar/R constriction site in AQP<sub>1</sub>, significantly higher than those in the PMFs obtained with US.

### 7.2.3 Structural fluctuations strongly affect PMFs from 3D-RISM

Since 3D-RISM calculations are carried out using a static protein structure, it is not clear how the protein dynamics, which are explicitly taken into account in US simulations, might affect the computation of the PMF. To rule out that the PMFs obtained from 3D-RISM are biased by the particular structural conformation of the UT-B and AQP<sub>1</sub> crystals, we took 100 different structures from equilibrium MD simulations of UT-B and AQP<sub>1</sub>, and computed the PMFs for urea permeation from 3D-RISM for each structure (see gray lines in Figure 7.4 A/C). In addition, we computed the mean PMF by averaging the 100 resulting densities from 3D-RISM (see Equation 6.5 in Section 6.2.1). The mean PMF is represented by a black line in Figure 7.4 A/C, which exhibits differences of  $\sim 120$  and  $\sim 25$  kJ mol<sup>-1</sup>, for UT-B and AQP<sub>1</sub>, respectively, between the mean PMF and the PMF from a single structure. This is indicative of a strong influence on 3D-RISM from the protein structural fluctuations at narrow sites inside the pore. Besides, an energy barrier  $\Delta G^\ddagger \sim 17$  kJ mol<sup>-1</sup> in the mean PMF for the permeation of urea across AQP<sub>1</sub> implies a urea permeating channel, which is not consistent with experimental evidence [250].

Given the stark effect of protein structural fluctuations on 3D-RISM calculations, we aimed to rule out the possibility that the PMFs for urea permeation calculated from 3D-RISM did not reflect the urea binding sites in UT-B because the channel lumen structure was not adapted to the presence of the solute. Therefore, we computed the PMF from 3D-RISM for each of the 100 structures taken from an umbrella window simulation, in which urea was bound to UT-B at  $z = 0.3$  nm (Figure 7.4 B, gray and black lines). The expected energy minimum at  $z = 0.3$  nm is absent in all PMFs, indicating that 3D-RISM could not identify the urea binding site in UT-B even when the pore structure has adapted to a urea molecule.

### 7.2.4 Alternative closure relations do not improve PMF calculations from 3D-RISM

3D-RISM solutions can be converged via several closure relations. In this work, we mainly used the KH closure, but to rule out the effect of the closure relation, we also tested the chained PSE-n closure as suggested by other authors [215]. Figure 7.5 presents the PMFs for the permeation of water, NH<sub>3</sub>, urea, O<sub>2</sub> and methanol across UT-B and AQP<sub>1</sub> calculated from 3D-RISM using different closure relations to

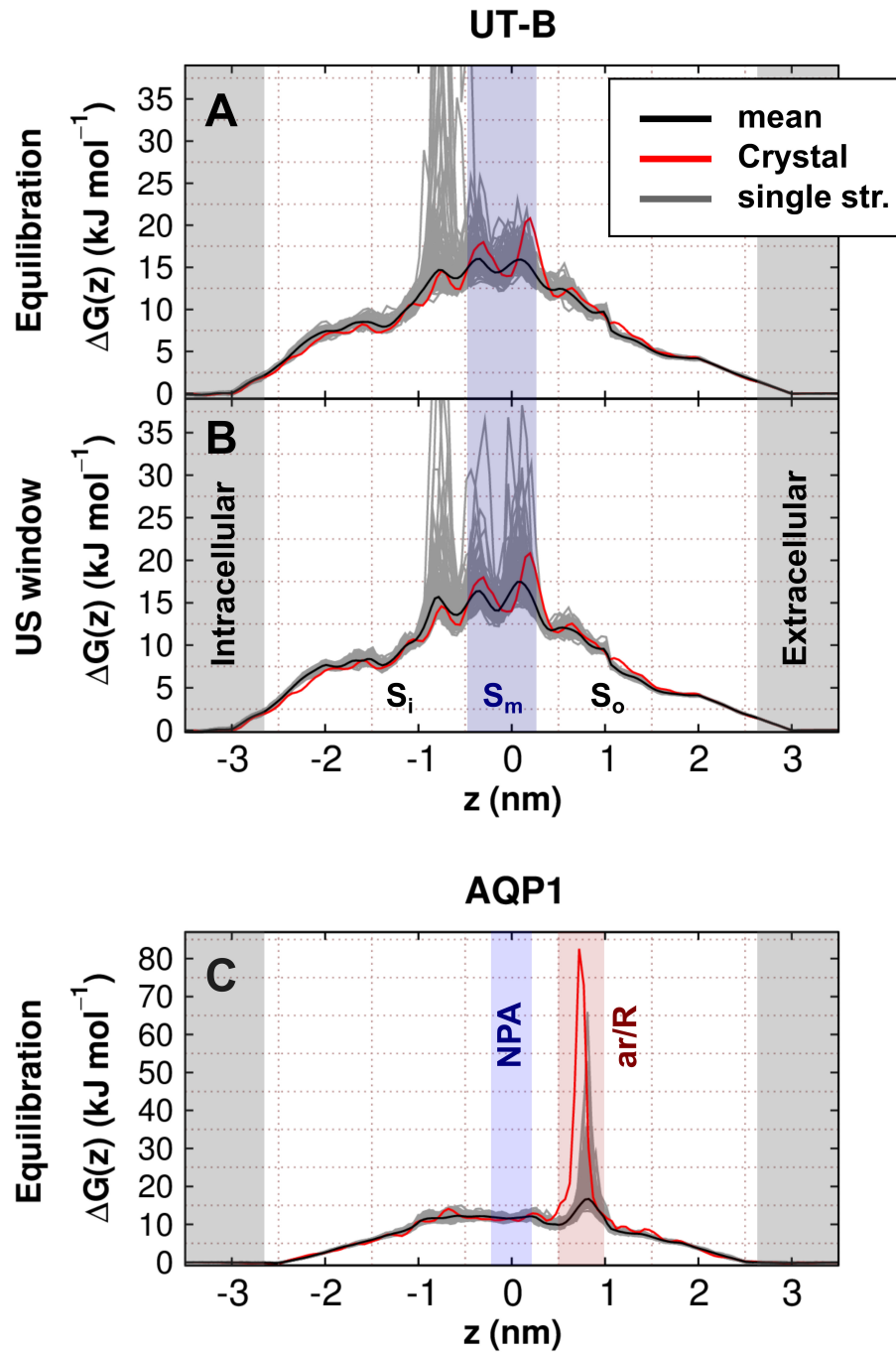


Figure 7.4: On the influence of structural fluctuations on PMFs from 3D-RISM. PMFs  $\Delta G(z)$  of urea permeation calculated from 3D-RISM across (A/B) UT-B and (C) AQP1 over 100 structures taken either from an equilibrium simulation or the umbrella window in which urea is at the binding site of UT-B (gray). The mean PMF of the respective 100 structures is represented by a black line, and the PMF obtained from the crystal structure by a red line. The  $S_i$ ,  $S_m$ , and  $S_o$  regions are highlighted for UT-B, as well as the NPA motif and the ar/R region by colored bars. The peaks of the single structure PMFs in A and B were removed for clarity.

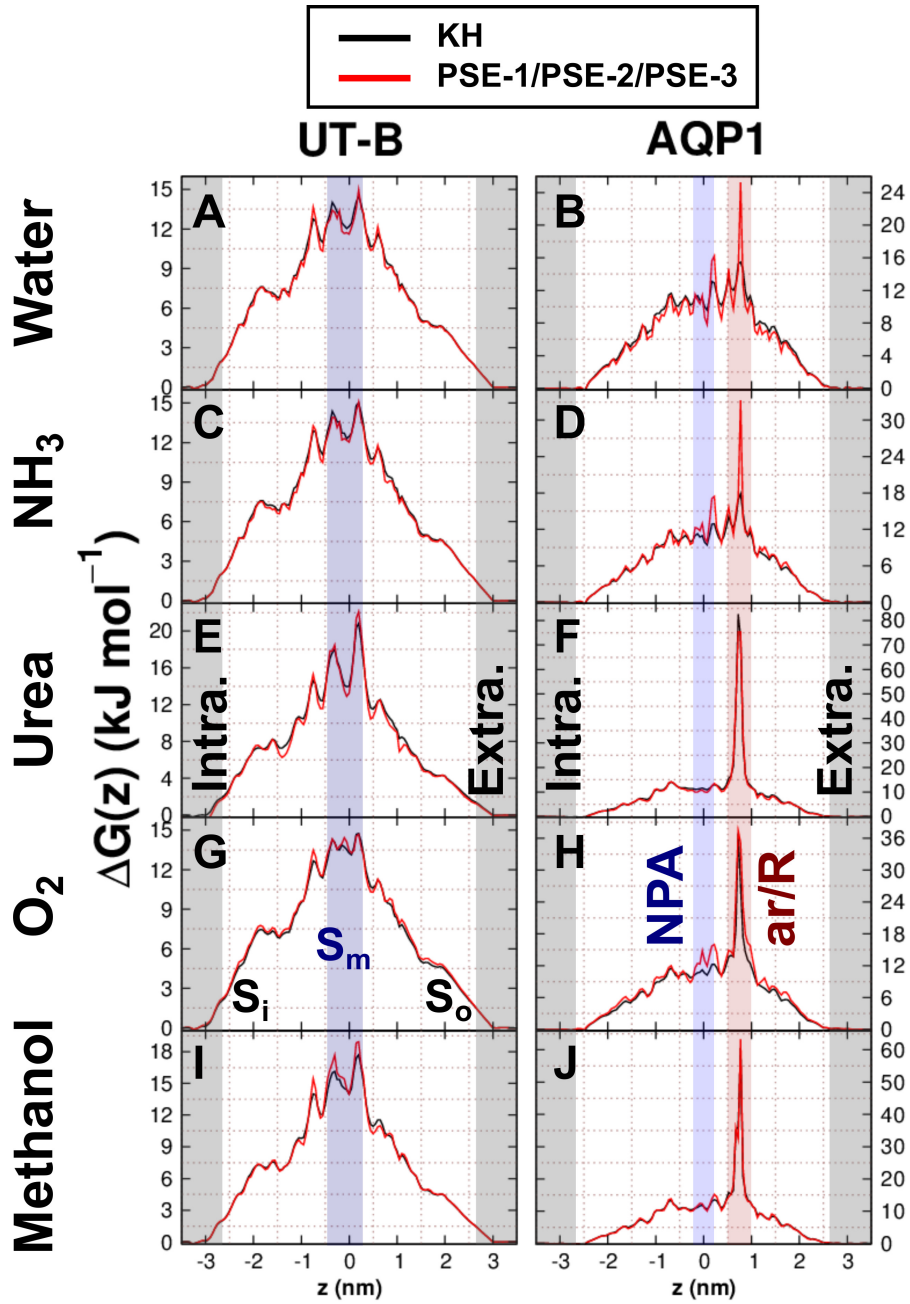


Figure 7.5: PMFs from 3D-RISM with KH and PSE- $n$  closures. PMFs  $\Delta G(z)$  for the permeation of water (A/B),  $\text{NH}_3$  (C/D), urea (E/F),  $\text{O}_2$  (G/H), and methanol (I/J) calculated with 3D-RISM across (A/C/E/G/I) UT-B, and (B/D/F/H/J) AQP1 using the KH closure and the chained PSE-1/PSE-2/PSE-3 closure. The  $S_i$ ,  $S_m$ , and  $S_o$  regions are highlighted for UT-B, as well as the NPA motif and the ar/R region for AQP1.

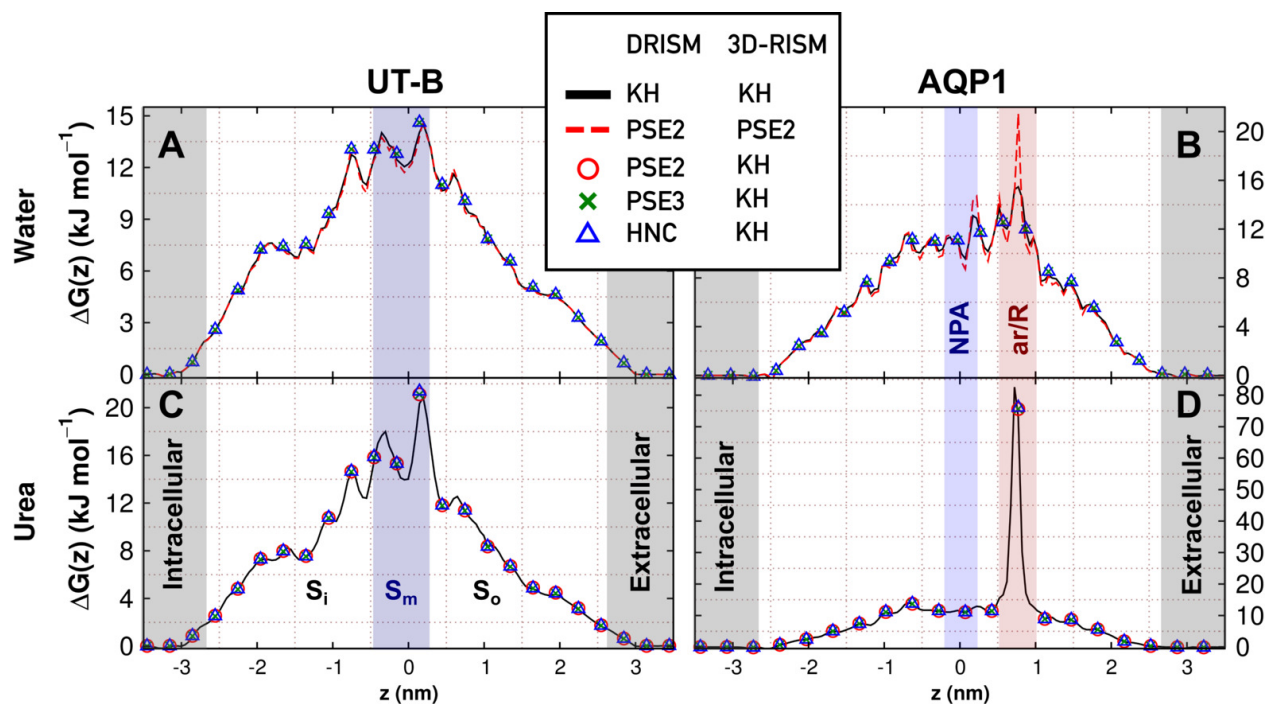


Figure 7.6: PMFs from 3D-RISM with KH, PSE- $n$ , and HNC are used for DRISM. PMFs  $\Delta G(z)$  for water (A/B), and urea (C/D) permeation calculated with either 3D-RISM-KH or 3D-RISM-PSE-2 (only red dashed line) across (A/C) UT-B, and (B/D) AQP<sub>1</sub>. The KH closure was used to solve DRISM equations, as represented by the solid line. The circles, crosses or triangles represent PSE-2, PSE-3 and HNC closures, respectively.

converge the solution, namely, KH closure, the chained PSE-1/PSE-2 closure, or the chained PSE-1/PSE-2/PSE-3 closure. Overall, the PMFs computed from 3D-RISM with either closure show no significant differences. Exceptions are the PMFs for water and NH<sub>3</sub> at the ar/R site in AQP<sub>1</sub>, in which differences between  $\sim 10$  and  $\sim 15$  kJ mol<sup>-1</sup>, respectively, are observed among different closure relations.

To solve the 3D-RISM equations, the solution to DRISM equations is required, in which a closure relation is also used (see Section 5.4). Therefore, we tested the effect of different closures applied to converge DRISM solution on PMF calculation from 3D-RISM-KH. Figure 7.6 shows the PMFs for water and urea permeation across UT-B and AQP<sub>1</sub> computed from 3D-RISM, in which the solutions for DRISM have been converged with different closures, namely, KH, PSE- $n$ , and HNC (see Section 6.2). These PMFs do not exhibit any significant differences, suggesting that the different closures applied to solve DRISM equations do not improve the DRISM-KH solutions.

Overall, these results suggest that neither the chained PSE- $n$  closure, nor the HNC closure, improve the performance of 3D-RISM-KH among the computation of PMFs for solute permeation across protein channels.

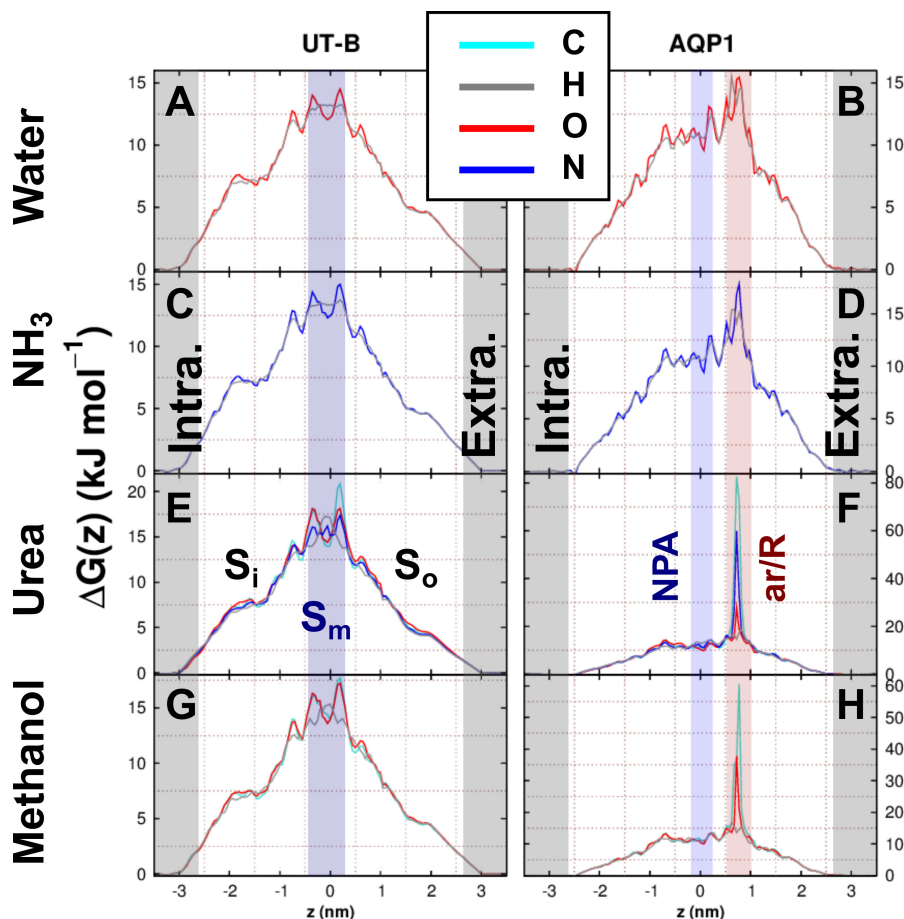


Figure 7.7: PMFs from 3D-RISM using different solvent sites. PMFs  $\Delta G(z)$  for the permeation of water (A/B), NH<sub>3</sub> (C/D), urea (E/F), and methanol (G/H) calculated with 3D-RISM across (A/C/E/G) UT-B, and (B/D/F/H) AQP1 using different solvent sites.

### 7.2.5 Different solvent sites in 3D-RISM yield similar PMFs

Since the PMFs calculated from 3D-RISM are derived from the 3-dimensional distribution function of a certain solvent-site among many possible, one could argue that the choice of that certain solvent-site may be determinant in the subsequent PMF. To rule out this possibility, we further calculated the PMFs of solute permeation across UT-B and AQP1 using every possible solvent site. The results, presented in Figure 7.7, indicate that in spite of the different solvent-site considered to calculate the PMFs, none of them show any of the features present in the PMFs computed from US simulations. Hence, the election of the solvent-site density distribution, upon which the PMF is calculated, does not seem to be of critical relevance.

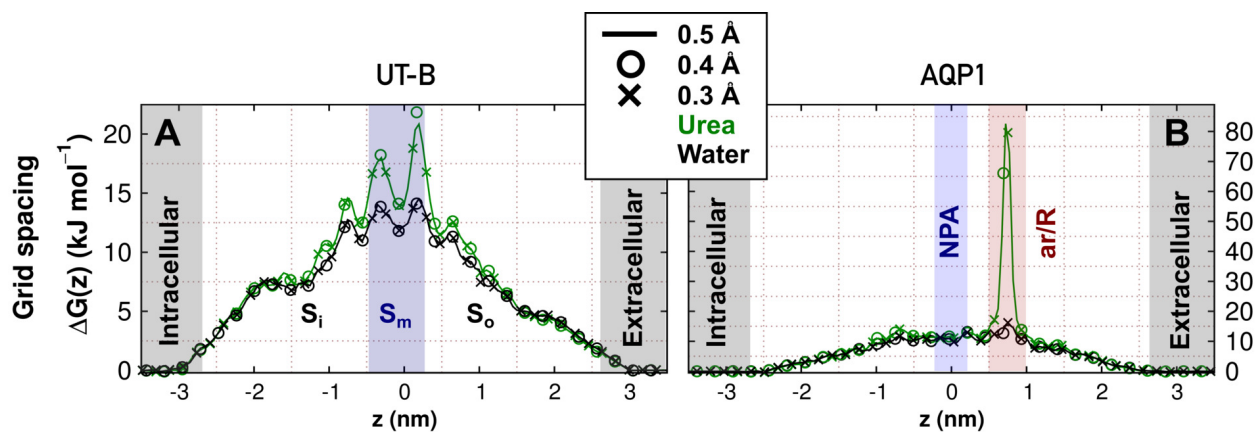


Figure 7.8: PMFs from 3D-RISM using different grid spacings. PMFs  $\Delta G(z)$  for the permeation of water and urea calculated with 3D-RISM across (A) UT-B, and (B) AQP1 using grid spacings of 0.3Å, 0.4Å, and 0.5Å.

### 7.2.6 More discrete grid spacing does not change the PMFs from 3D-RISM

To test the effect of the grid spacing used in 3D-RISM calculations, we computed the PMFs for water and urea across UT-B and AQP1 using grid spacings of 0.3Å, 0.4Å, and 0.5Å. These PMFs are shown in Figure 7.8. Despite the subtle differences, choosing a different grid size does not improve the PMFs computed from 3D-RISM using a 0.5Å grid spacing and, consequently, nor they improve the PMFs calculated from US simulations.

### 7.2.7 Urea COM vs. C-atom and the LJ parameters on H-atoms do not affect PMFs from US

It is important to stress that the PMFs calculated from US simulations are not strictly equivalent to the PMFs computed from 3D-RISM for two reasons: i) whilst the PMFs computed from US simulations usually take the solute COM as the reaction coordinate, the PMF calculated from 3D-RISM takes the C-atom of urea as the referent solvent site. ii) The H-atoms do not have LJ interactions in the US simulations carried out in this work, whereas the H-atoms do have LJ interactions in 3D-RISM calculations. Therefore, to test the impact of these differences on PMF calculation, we computed the PMFs for urea permeation across UT-B from US simulations in which the H-atoms of urea molecules have the same LJ parameters as in 3D-RISM calculations, either using the COM of urea, or its C-atom as the reaction coordinate reference. The resulting PMFs, which are shown in Figure 7.9 (red and green curves), exhibit slightly different profiles as compared with the conventional PMF computed from US (black curve), but still present the energy minima that correspond to urea binding sites. This suggests that neither

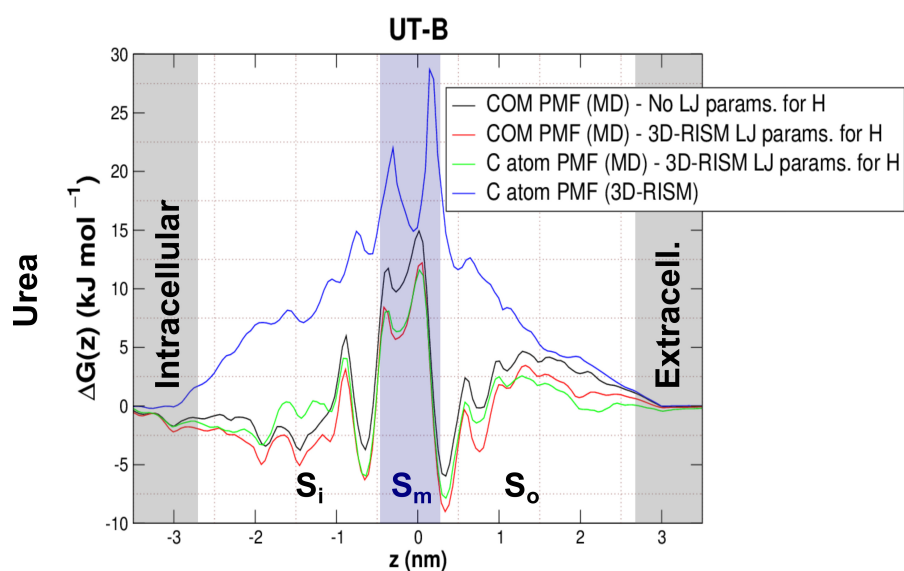


Figure 7.9: PMFs from US using the C-atom of urea and LJ parameters for H-atoms. PMFs  $\Delta G(z)$  for urea permeation calculated from US simulations across UT-B, either using urea COM or C-atom as the reference for the reaction coordinate, and different LJ parameters for H-atoms. The blue curve represents the typical PMF for urea computed from 3D-RISM, whereas the black curve shows the typical PMF calculated from US, using a urea molecule with no LJ parameters for H-atoms. The red curve shows a PMF calculated using a urea molecule with the LJ parameters for H-atoms used typically in the 3D-RISM calculations, and the green curve shows the PMF calculated from US using the urea C-atom as the reference for the reaction coordinate, using a urea molecule with the 3D-RISM conventional LJ parameters for H-atoms.

the LJ parameters for the H-atoms used in 3D-RISM calculations, nor the choice of the reaction coordinate reference (COM or C-atom) explain the large differences observed between the PMFs calculated from US and the one calculated from 3D-RISM (blue curve).



---

## ASSIGNING THE CRYSTALLOGRAPHIC ELECTRON DENSITIES IN FLUC

---

**M**ANY PROTEINS bind ions and water at well-defined binding sites, where they play important roles in the functions and structural stability of the proteins. For instance, water and ions may contribute to the active site of enzymes, water molecules may function as proton donor or acceptors, and ions frequently bind into selectivity filters of ion channels. Hence, understanding protein function and structure at the atomic level requires precise knowledge of binding sites for water and ions. Such binding sites may be revealed by the appearance of small isolated electron densities in the density maps derived from protein crystallography or cryo-electron microscopy. However, the assignment of such densities to water, ions, or other small molecules is often far from obvious, and instead requires substantial human supervision and experience. Here, we argue that MD simulations and free energy calculations may guide the assignment of isolated electron densities. As a test case, we analyze isolated electron densities in the crystal structure of the fluoride channel Fluc. The simulations suggest, in the light of the applied force-fields, that four densities previously assigned to fluoride ions more likely represent four water molecules.

### 8.1 WATER IS MORE STABLE THAN $F^-$ INSIDE FLUC-BPE IN FREE MD SIMULATIONS

As explained in Chapter 3, four fluoride ions were assigned to four disconnected electron densities found in the Fluc-Bpe crystal at F82 and F85 sites (Figure 3.2) [85]. To test the stability of fluoride ions at F82 and F85, we performed MD simulations in crystallographic and physiological conditions, i. e. with and without pore-blocking L2 monobodies, respectively (see details in Section 6.5). Accordingly, we ran 10 replicas in which  $F^-$  ions started at F82 and F85 and Fluc was blocked by L2 monobodies (80 ns each), and 20 replicas in which Fluc was free of L2 monobodies (40 ns each). Analogous simulations were carried out with water molecules starting at F82 and F85. Figure 8.1 shows typical examples of the positions in time of  $F^-$  and water molecules along the Fluc z-axis in each system. In addition,

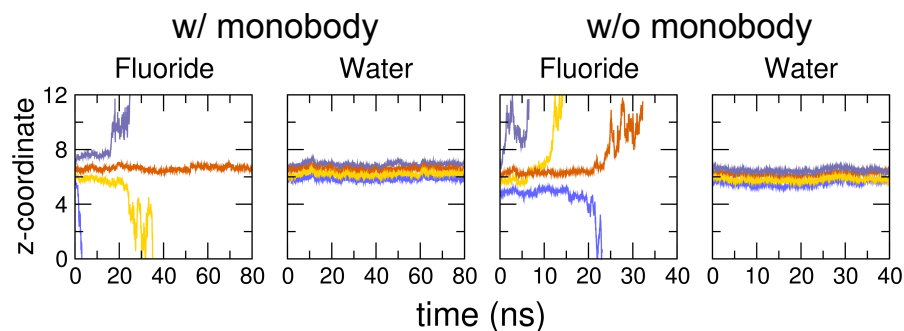


Figure 8.1:  $F^-$  and water positions at the F82 and F85 sites in Fluc. Position change in  $z$ -direction (normal to membrane) in MD simulations. (A)  $F^-$  position inside Fluc with pore-blocking L2 monobodies. (B) Water molecule position in Fluc with pore-blocking L2 monobodies. (C)  $F^-$  position inside Fluc without monobodies. (D) Water molecule position in Fluc without monobodies. Orange curves represent the molecules starting at F82, and blue curves represent the molecules starting at F85. Different shades of orange and blue indicate starting positions in the two Fluc homodimers.

to test the influence of the force-field in these simulations, we carried out 10 additional replicas using the CHARMM36 force-field in which Fluc was free of monobodies, and another 10 similar replicas using the polarizable Drude force-field. The plots for all simulations are shown in Appendix B.

To quantify the results from the above-mentioned MD simulations, we calculated the maximum likelihood estimates for the lifetime  $\tau_{ml}$  of  $F^-$  and water at F82 and F85, which are summarized in Table 8.1 (see Section 6.7). The estimates indicate that the lifetime of water is one order of magnitude longer than that of  $F^-$ . When Fluc-Bpe was blocked by L2 monobodies, not a single water molecule escaped the channel within a total of 1600 ns whereas, in contrast,  $F^-$  escaped within 181 ns and 25 ns from F82 and F85, respectively. When Fluc-Bpe was in the absence of monobodies, water presented slightly reduced lifetimes of 1600 ns and 1595 ns at F82 and F85, respectively.  $F^-$  lifetimes were even shorter.  $F^-$  escaped within 108 ns and 3 ns from F82 and F85, respectively. The simulations carried out using the CHARMM36 or the polarizable Drude force-fields instead of Amber ff99SB\*-ILDN yielded similar results in the absence of pore-blocking monobodies (Table 8.1, numbers in parentheses). Overall, the simulations showed that water molecules at F82 and F85 were more stable as compared with  $F^-$  ions, regardless the presence or absence of monobodies, and regardless the applied force-field, suggesting that the disconnected electron densities at F82 and F85 are rather ordered water molecules and not  $F^-$ .

Table 8.1: Maximum-likelihood estimates  $\tau_{\text{ml}}$  (nanoseconds) for the lifetime of  $\text{F}^-$  and water molecules at Fluc-Bpe F82 and F85 sites.

	w/ monobody fluoride	w/ monobody water	w/o monobody fluoride	w/o monobody water
F82	181	$\gtrsim 1600$	108 (91; 44)	$\gtrsim 1600$
F85	25	$\gtrsim 1600$	3(3;1)	1595

The numbers not in parentheses have been calculated with the Amber ff99SB\*-ILDN force-field, and the numbers in parentheses have been calculated applying the CHARMM36 force-field and the polarizable CHARMM-Drude force-field, respectively (first and second number in parentheses).

## 8.2 DIFFERENCE IN $\Delta\Delta G_{\text{BIND}}$ BETWEEN $\text{F}^-$ AND WATER TO FLUC FAVORS WATER OVER $\text{F}^-$

In the previous section, free MD simulation results indicate that water molecules are more stable than  $\text{F}^-$  ions at the Fluc-Bpe F82 and F85 binding sites. To rationalize these results, we used Thermodynamic Integration (TI) to compute the relative binding free energy differences  $\Delta\Delta G_{\text{bind}}$  between  $\text{F}^-$  and water at F82 and F85 (see Section 6.6).

Figure 8.2A shows  $\Delta\Delta G_{\text{bind}}$  as blue and green bars, in which a negative value is equivalent to a higher probability of water binding. When Fluc-Bpe is blocked by L2 monobodies, and thus preventing the water molecules outside the channel from solvating the molecules at F82 and F85 (see Figure 8.3A/B/E/F), the free energy differences at F82 and F85 are  $\Delta\Delta G_{\text{bind}} \approx -235 \text{ kJ mol}^{-1}$ , indicating a much more favorable binding for water at those sites as compared with  $\text{F}^-$  (Figure 8.2A, solid blue/green bars). This large difference is explained by the large dehydration free energy for  $\text{F}^-$  of  $\sim 500 \text{ kJ mol}^{-1}$ , which is only partially compensated by  $\text{F}^-$ -protein interactions [251]. When Fluc-Bpe was free of monobodies, the calculated  $\Delta\Delta G_{\text{bind}}$  values were less negative, although still indicating a higher probability for water molecules over  $\text{F}^-$  of binding at F82 and F85 (Figure 8.2A and Figure 8.3G/H). This trend is explained by the partial solvation of the  $\text{F}^-$  restrained at F82 and F85 which, in the absence of monobodies, drag water molecules from bulk thereby stabilizing  $\text{F}^-$  (see Figure 8.3C/D). Thus, the calculated  $\Delta\Delta G_{\text{bind}}$  are consistent with the results from Section 8.1, indicating a more favorable binding of water as compared with  $\text{F}^-$  to F82 and F85, regardless of the presence or absence of pore-blocking monobodies.

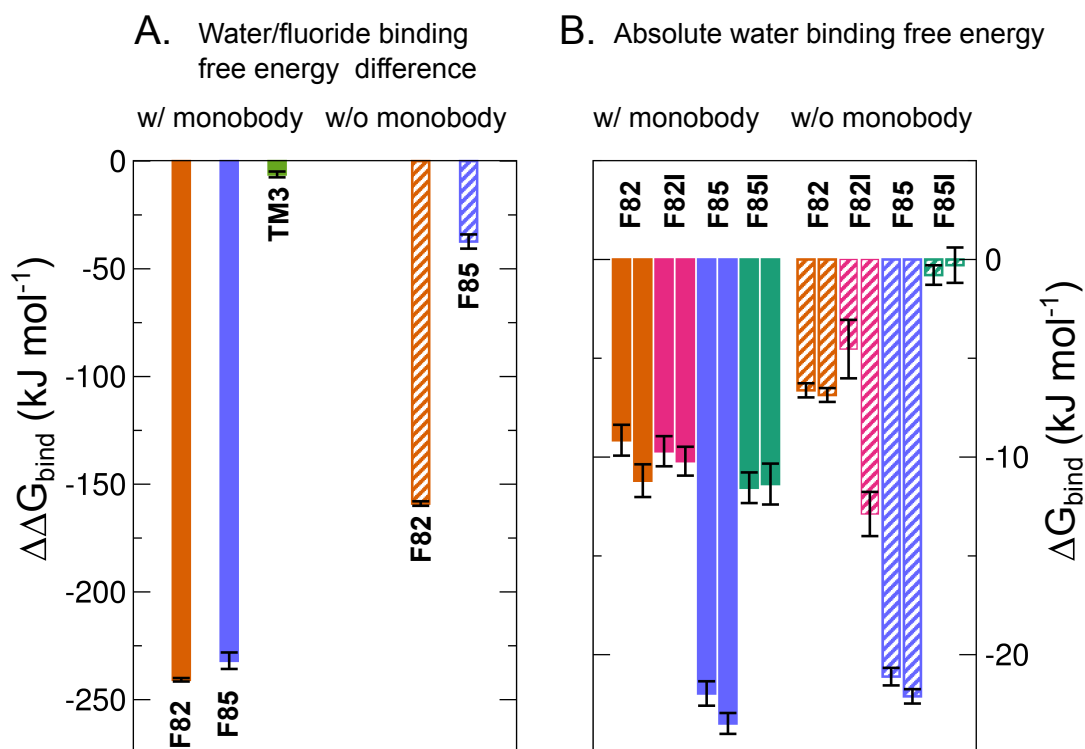


Figure 8.2: **Summary of free-energy calculations.** (A) Difference  $\Delta\Delta G_{\text{bind}}$  of the binding free energy of water versus an ion, defined such that a negative  $\Delta\Delta G_{\text{bind}}$  implies a more favorable binding free energy for water. Color code as follows. Orange:  $\text{F}^-/\text{water}$  bound to F82 site; blue:  $\text{F}^-/\text{water}$  bound to F85 site; olive:  $\text{Na}^+/\text{water}$  bound to TM3 site. Solid and dashed bars show results computed in the presence and absence of the L2 pore-blocking monobodies. (B) Absolute binding free energies  $\Delta G_{\text{bind}}$  of water to F82 (orange) and F85 (blue) sites in the WT channel, and to the F82I (magenta) and F85I (dark teal) sites in the mutant channel. The same filled/dashed bar scheme is followed. The two values for each site correspond to the two Fluc monomers.

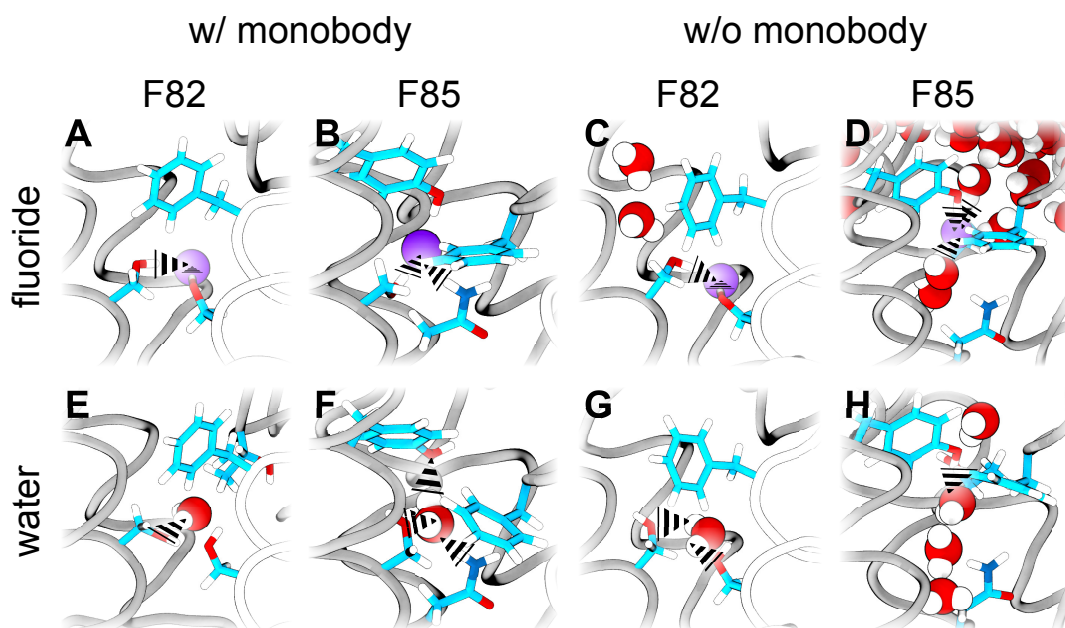


Figure 8.3: **Water and  $\text{F}^-$  at F82 and F85 sites, taken from snapshots of the relative binding free energy simulations.** Several representative snapshots taken from the alchemical transformation simulations are depicted. Fluoride (A,B,C,D) is shown as purple spheres and water (E,F,G,H) as red (oxygen) and white (hydrogen) spheres. Some of the amino-acid side-chains in the F82 (A,C,E,G) and F85 (B,D,F,H) sites are depicted in colored sticks. Some of the hydrogen bonds established between the protein and the water/fluoride are highlighted with black dashed triangles.

Table 8.2: Absolute binding free energy  $\Delta G_{\text{bind}}$  ( $\text{kJ mol}^{-1}$ ) of water at F82, F85, F82I, and F85I sites.

	F82	F85	F82I	F85I
w/ L2 monobodies	$-11.2 \pm 0.8$	$-23.5 \pm 0.5$	$-10.2 \pm 0.7$	$-11.4 \pm 1$
	$-9.1 \pm 0.8$	$-22 \pm 0.6$	$-9.7 \pm 0.8$	$-11.6 \pm 0.8$
w/o L2 monobodies	$-6.6 \pm 0.4$	$-22.1 \pm 0.4$	$-12.9 \pm 1.1$	$-0.3 \pm 0.9$
	$-6.9 \pm 0.4$	$-21.1 \pm 0.4$	$-4.5 \pm 1.5$	$-0.8 \pm 0.5$

### 8.3 ABSOLUTE BINDING FREE ENERGY $\Delta G_{\text{BIND}}$ OF WATER TO FLUC

The  $\Delta\Delta G_{\text{bind}}$  calculations between water and  $\text{F}^-$  favour the binding of water over  $\text{F}^-$  to the F82 and F85 sites. However, since these numbers are mostly affected by the large  $\text{F}^-$  dehydration penalty, they do not prove that water molecules, *per se*, are able to bind to F82 and F85, which is needed to justify the assignment of water molecules to the disconnected electron densities at F82 and F85. Therefore, we calculated the absolute binding free energy  $\Delta G_{\text{bind}}$  of water to F82 and F85, i. e. the free energy for translocating a water molecule from bulk solvent into the F82 and F85 sites, again either in presence or in absence of pore-blocking monobodies (see Section 6.6).

Figure 8.2B and Table 8.2 show  $\Delta G_{\text{bind}}$  of water at F82 and F85 in WT Fluc-Bpe. The negative  $\Delta G_{\text{bind}}$  in both the presence and absence of monobodies (solid and shaded bars, respectively) indicate a favourable water binding to these sites. The F85 site offers several interaction sites as hydrogen bond donors and acceptors such as Ser180, Tyr104, and Asn43, thus yielding large negative  $\Delta G_{\text{bind}}$  of  $\approx -20 \text{ kJ mol}^{-1}$ . The F82 site offers fewer interaction sites, namely the hydroxyl group of Ser112, and the backbone carbonyl of Asn43 the main interacting partners for water, which yields less negative  $\Delta G_{\text{bind}}$ . Overall, these values are consistent with the binding of water to F82 and F85.

### 8.4 RATIONALIZING THE DISAPPEARANCE OF ELECTRON DENSITIES IN THE F82I/F85I MUTANT.

The loss of the disconnected electron densities associated to the F80I and F83 sites in the non-functional F80I/F83I Fluc-Ecl mutant (analogues of F82 and F85 sites in Fluc-Bpe), was interpreted by Last *et al.* as an indication of the functional relevance of these  $\text{F}^-$  binding sites [97]. Since our results have hitherto suggested that F82 and F85 are not  $\text{F}^-$  binding sites, but rather water binding sites, we tested whether the assignment of water to the electron densities is compatible to the disappearance of such densities in the mutant channel. Hence, we fur-

ther calculated the absolute free energy of binding  $\Delta G_{\text{bind}}$  of water to the F82I and F85I sites in Fluc-Bpe.

Figure 8.2B (magenta and dark green bars) shows the binding free energies  $\Delta G_{\text{bind}}$  of water to F82I and F85I remain mostly negative upon mutation, indicating that water would occupy the mutant binding sites as well. The only exception is the F85I site when monobodies are absent, in which, as soon as the binding water molecule is moved to bulk, it is replaced by a water molecule from the bulk, yielding a  $\Delta G_{\text{bind}} \sim 0$ . These results may seem incompatible with the disappearance of electron densities in the mutant. However, the loss of electron densities might also be explained by an increased disorder of bound water molecules in the crystal, in consequence of the increased volume of the cavity in the Phe→Ile mutant. We tested this hypothesis by running free MD simulations, for 80 ns each, of water at F82, F85, F82I, and F85I, and we compared the disorder of the water molecules between the WT and the mutant. Figure 8.4A/B shows an overlay of the positions of water oxygen atoms taken from  $\sim 2000$  MD snapshots of WT and mutant channels. More quantitatively, the volume sampled by water was calculated from the Root Mean Squared Fluctuations (RMSF) of water oxygen atoms at F82 and F85 in each simulation, taking the mean oxygen position as the reference. The average RMSF of oxygen atoms in WT Fluc-Bpe is 0.032 nm, and 0.038 nm at F82 and F85, respectively, which is consistent with highly ordered water molecules. In contrast, the average RMSF of oxygen atoms in the mutant increases to 0.066 nm, and 0.063 nm at F82I and F85I, respectively, which indicates that water molecules are significantly more dispersed in the mutant as compared with the WT channel. These findings rationalize the disappearance of the electron densities in the mutants not as a consequence of loss of water binding, but due to increased disorder of bound water molecules in the mutant crystal.

### 8.5 TM3 SITE IS STABILIZED BY $\text{Na}^+$

Stockbridge *et al.* assigned  $\text{Na}^+$  to the disconnected electron density at the TM3 site in Fluc-Bpe crystal, based on its coordination by four main chain carbonyl oxygen atoms (Figure 3.2C, blue sphere) [85]. To test whether MD simulations support this assignment, we computed the relative binding free energy difference  $\Delta\Delta G_{\text{bind}}$  between  $\text{Na}^+$  and water at TM3. Figure 8.2A shows a  $\Delta\Delta G_{\text{bind}} = -6.2 \pm 1.4 \text{ kJ mol}^{-1}$ , which suggests slightly more favourable binding of water over  $\text{Na}^+$ . However, it is important to stress that the hydration free energy of  $\text{Na}^+$  is  $\sim -350 \text{ kJ mol}^{-1}$ , and hence the resulting  $\Delta\Delta G_{\text{bind}}$  is a small number derived as the difference between two large numbers, which is certainly not within the accuracy of the force-fields. This leads to the conclusion that free energy calculations, in this case, are insufficient to assign the isolated electron density.

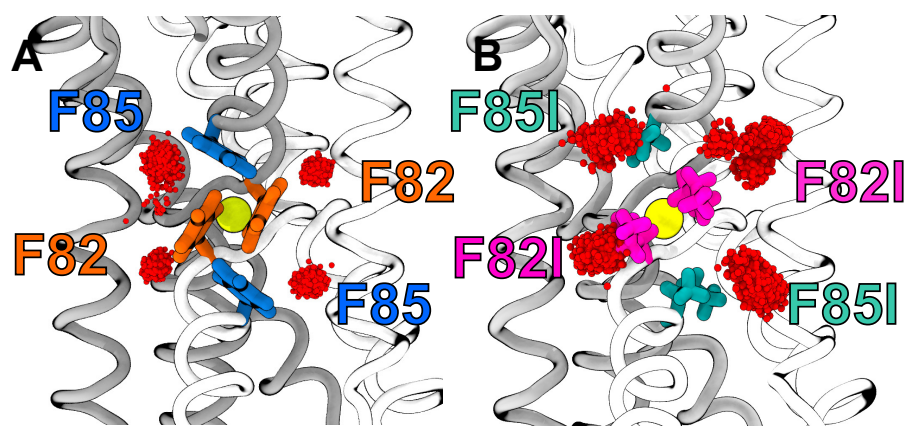


Figure 8.4: **On the disappearance of electron densities at the F82I and F85I in the mutant crystals.** Water oxygen atoms (red spheres) positions in the 2000 snapshots taken from 80-nanosecond simulations. (A) WT Fluc channel is shown in grey and white ribbons. The F82 and F85 binding sites are highlighted in both monomers. (B) Mutant Fluc channel is depicted as before. Here, the F82I and F85I mutated residues are highlighted.

Therefore, we also tested the structural stability of the TM3 site in free MD simulations with either  $\text{Na}^+$  or water coordinated at TM3. Figure 8.5A/B shows the positions of G77 and T80 residues forming the TM3 site, which were taken from 10 ns free MD simulations (thin sticks), as compared with the positions of G77 and T80 residues in the crystal (thick sticks). Both G77 and T80 residues remain close to the crystal structure when  $\text{Na}^+$  is bound at TM3. However, when water binds to TM3, its structure is more flexible and at least one G77 residue significantly changes its orientation as compared with the crystal, possibly because water cannot establish four simultaneous hydrogen bonds with the carbonyl groups [85]. Hence, whereas free energy calculations yielded inconclusive results, free MD simulations confirm  $\text{Na}^+$  as the most plausible species at the TM3 site.

## 8.6 THE ROLE OF THE FORCE-FIELD IN MODELING ION-PHENYLALANINE INTERACTIONS

We further tested whether force-field uncertainties may influence our conclusions on the nature of the isolated electron densities in Fluc-Bpe. To this end, we computed the relaxed Potential Energy Surface (PES) between  $\text{F}^-$  and benzene —used as a model for the Phe side chain— using quantum-chemical calculations, and we compared them with results from additive and polarizable force-fields (see Section 6.8). The PESs in Figure 8.6B reflect a strong interaction between  $\text{F}^-$  and the quadrupole of benzene, revealing a potential energy of binding in vacuum of  $\sim -50 \text{ kJ mol}^{-1}$ . The fluoride-phenyl interactions modelled by the additive Amber ff99SB\*-ILDN force-field with



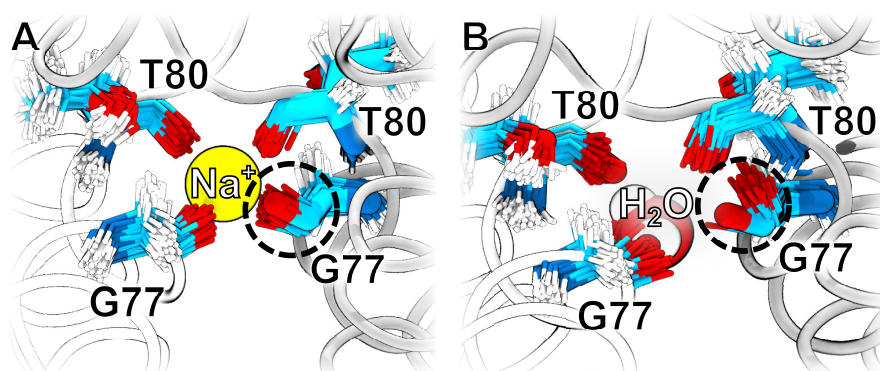


Figure 8.5: **G77 and T80 residue positions in MD simulations.** (A) Positions of G77 and T80 residues in a 10 ns MD simulation in which  $\text{Na}^+$  is bound to TM3. The positions of G77 and T80 residues are shown as thin sticks in which oxygen is colored in red, hydrogen in white, nitrogen in blue, and carbon in cyan. The positions of the G77 and T80 residues in the crystal structure are depicted in thick sticks following the same color scheme. The protein secondary structure is showed in white and gray ribbons. (B) Positions of G77 and T80 residues in a 10 ns MD simulation in which a water molecule is bound to TM3. The G77 residue that adopts a systematically different orientation in the simulation as compared with the crystal structure is highlighted in a dashed circle.

the ion parameters by Joung & Cheatham III are significantly underestimated, suggesting that the relative binding free energy difference  $\Delta\Delta G_{\text{bind}}$  between  $\text{F}^-$  and water, showed in Figure 8.2A, is overestimated by  $\approx 40 \text{ kJ mol}^{-1}$  in favour of water. It is important to highlight that such correction would not lead to different qualitative conclusions, given the  $\Delta\Delta G_{\text{bind}} \sim -235 \text{ kJ mol}^{-1}$  yielded by the additive force-field in the crystallographic conformation. Besides, the fluoride-benzene interactions modelled by the polarizable CHARMM-Drude force-field were stronger and more consistent with QM calculations. However, despite the stronger  $\text{F}^-$ -benzene interactions modelled by CHARMM-Drude, the estimated lifetime of  $\text{F}^-$  at F82 and F85 in simulations carried out with that same force-field was even shorter as compared with Amber ff99SB\*-ILDN simulations (Table 8.1, second numbers in parentheses). This result indicates that the stability of  $\text{F}^-$  at these binding sites is not critically dependent on the fluoride-phenyl interactions.

Additionally, we computed the relaxed PESs between  $\text{Cl}^-$  and benzene via QM calculations and compared them with results from additive and polarizable force-fields. Figure 8.6C shows that the chloride-phenyl interaction modelled by the B3LYP/ma-def2-TZVP—which uses the most accurate basis set among the ones tested in this study—yields a potential energy of binding in vacuum of  $\sim -23 \text{ kJ mol}^{-1}$ , whilst the same interaction modelled by the additive force-field is

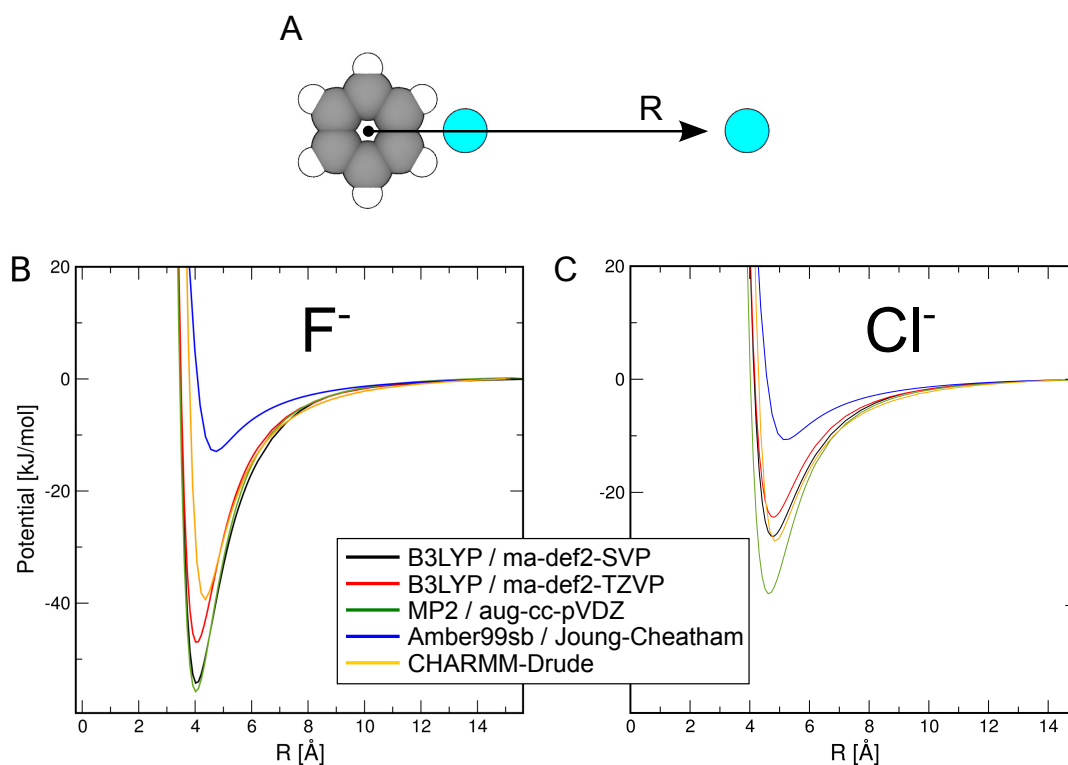


Figure 8.6: **Relaxed PESs of ion-benzene interactions.** (A) Illustration of the coordinate  $R$  as the distance of ion (cyan sphere) from the center of mass of the benzene ring (green/white spheres), taken in the plane of the ring. The left cyan sphere indicates the position of the potential energy minimum. (B) Relaxed PESs of fluoride-benzene using QM calculations at different quantum levels and using different basis sets, as indicated in the legend (black, red, and green curves). (C) Relaxed PESs of chloride-benzene using QM calculations, as well as the additive Amber ff99SB\*-ILDN and the polarizable CHARMM-Drude force-field.

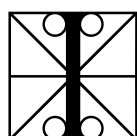
underestimated by  $\sim 10 \text{ kJ mol}^{-1}$  as compared with the B3LYP/ma-def2-TZVP result. Again, the results from the polarizable CHARMM-Drude force-field are in better agreement with the QM calculations. This calls into question the suitability of additive force-fields to unveil the selectivity mechanism of Fluc.



---

 THE PERMEATION MECHANISM IN FLUC-BPE
 

---



ION CHANNELS mediate the selective flux of ionic species across cell membranes. Their function and structure can be studied by a wide range of experimental techniques, such as electrophysiology and crystallography. Yet computational methods are essential to understand in atomic detail the permeation and selectivity mechanisms of ion channels. In this study we use MD and US simulations to analyze the structural and functional properties of the suggested permeation pathway in the fluoride channel Fluc-Bpe. Accordingly, each Fluc-Bpe homodimer presents two permeating pores. In turn, each pore can be identified by the two vertically aligned electron densities that were assigned to fluoride in the crystal structure. Our results indicate that i) the structure of the channel opens upon keeping the fluoride ions restrained at the F82 and F85 binding sites owing to water molecules entering the pore, and ii) that this open structure allows the efficient and non-selective permeation of water,  $F^-$ , and  $Cl^-$ .

### 9.1 WATER ENTERS FLUC-BPE WHEN $F^-$ IONS ARE RESTRAINED IN THE CHANNEL

The results shown in Chapter 8 leave open the question about the permeation mechanism in Fluc channels. With the aim to find a  $F^-$ -conducting state of Fluc, we carried out MD simulations of the Fluc-Bpe system, in the absence of pore-blocking monobodies, in which a restraining potential was applied to the fluoride ions at F82 and F85. We ran 20 replicas with the Amber ff99SB\*-ILDN force-field (80 ns each) and, to test the influence of the applied force-field, we ran another 10 replicas with the CHARMM36 force-field (80 ns each). Figure 9.1 shows the result of one such simulation, in which the water molecules in bulk enter the pores and solvate the restrained  $F^-$  ions. The trans-membrane helices TM2, TM3a, and TM3b move to accommodate the increasing number of solvent molecules that surround the ions. These additional water molecules occupy the lumen of the pore between the ions, transforming Fluc-Bpe into a water-filled channel.

More quantitatively, Figure 9.2 shows typical time traces of the running average number of water molecules inside Fluc-Bpe as com-

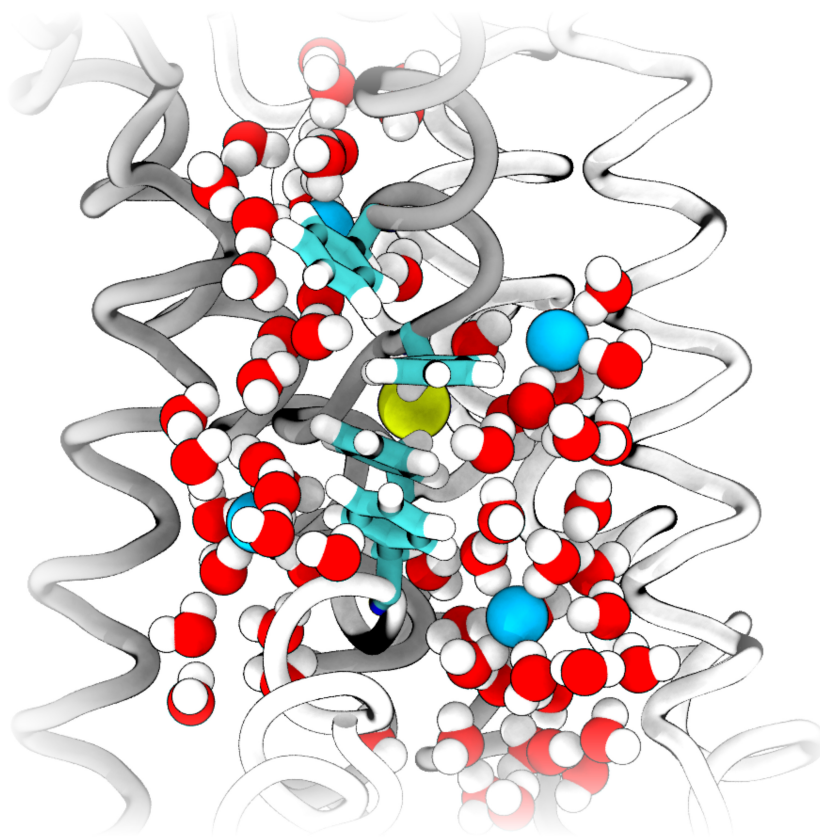


Figure 9.1: **The structure of Fluc-Bpe when the pores are filled with water molecules.** A single snapshot taken from one simulation in which Fluc-Bpe pores are *open* and filled with water molecules while the F<sup>-</sup> ions remain restrained at the crystallographic binding sites. The F<sup>-</sup> ions are represented as blue spheres, the water molecules as red/white spheres, and the F82 and F85 residues as thick sticks.

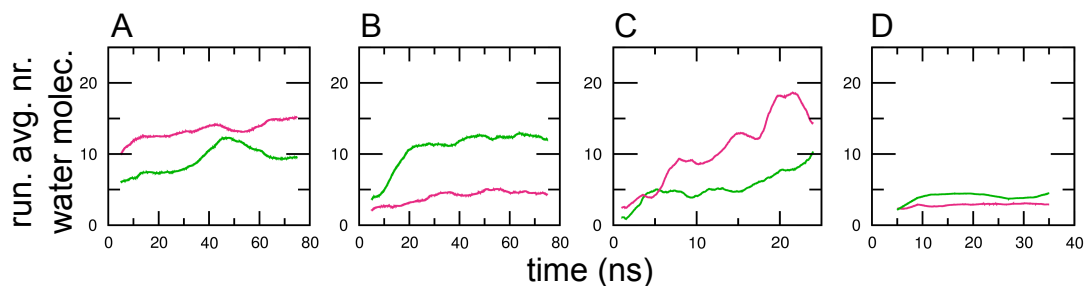


Figure 9.2: **Number of water molecules inside Fluc-Bpe.** Typical time traces of the running average number of water molecules inside Fluc-Bpe in the absence of pore-blocking monobodies. The running averages are taken from simulations with the (A) Amber ff99SB\*-ILDN force-field, (B) the CHARMM36 force-field, or (C) the CHARMM-Drude force-field, in which F<sup>-</sup> ions were restrained at the F82 and F85 binding sites, or (D) in which water molecules were bound to F82 and F85. Each color represents one of the pores.

Table 9.1: **Number of water molecules in Fluc-Bpe pores.**

Molec. at F82/F85	F <sup>-</sup> (Amber99SB)	F <sup>-</sup> (CHARMM36)	F <sup>-</sup> (CHARMM-Drude)	Water
avg. nr. of water molec. / pore	10.29	9.96	9.05	3.01
avg. RMSD (nm)	0.23	0.24	0.29	0.11

puted from the simulations. The number of water molecules inside each pore was calculated as the number of water-oxygen atoms within the volume of a cylinder. The geometric center of the cylinder is defined as the COM of the pore C<sub>α</sub> atoms, with radius  $r = 0.4$  nm and height  $h = 2$  nm. To compare these results with the number of molecules that enter Fluc-Bpe in the simulations in which water molecules are bound to the binding sites, we also computed the running averages in such simulations. The plots of running averages of each simulation can be seen in Appendix C. Overall, the results show that more water molecules enter Fluc-Bpe when F<sup>-</sup> ions are kept inside, as compared with the simulations in which four water molecules are bound to F82 and F85.

The results shown in Figure 9.2 and in Appendix C are summarized in Table 9.1. The results indicate that the restrained F<sup>-</sup> ions inside Fluc-Bpe drag  $\sim 10$  water molecules to each pore, whereas the two water molecules bound to F82 and F85 drag only  $\sim 1$  water molecule (the number in Table 9.1 already includes the two water molecules bound to Fluc-Bpe). In addition, we calculated the average Root Mean Squared Deviation (RMSD) between the backbone atoms of the last frame of the simulations and the crystal structure. With average RMSD values  $> 0.2$  nm, the Fluc-Bpe structure clearly deviates from the crystal, reflecting the deformation of TM2, TM3a, and

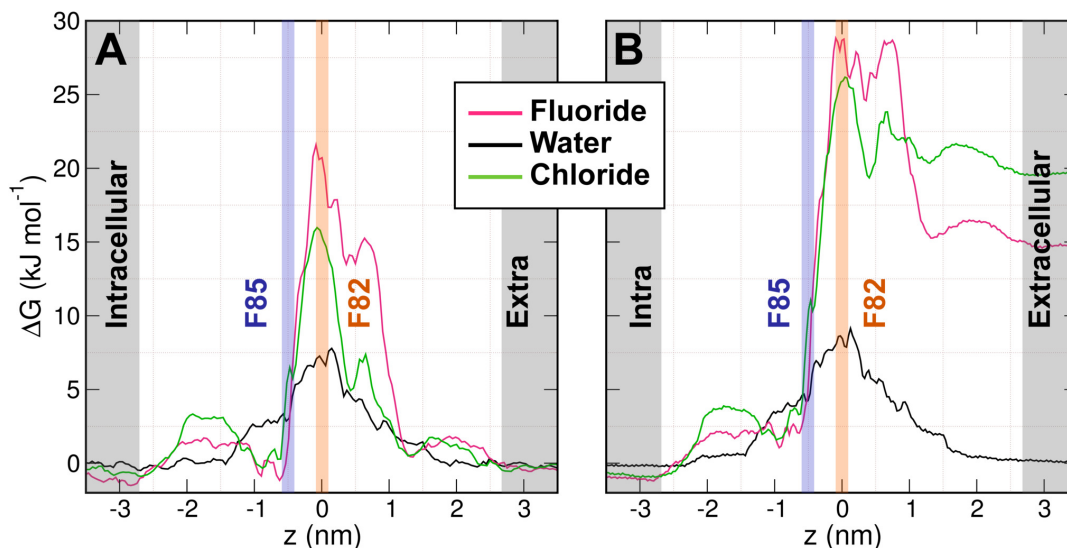


Figure 9.3: **PMFs of solute permeation across Fluc-Bpe.** PMFs  $\Delta G(z)$  for the permeation of  $F^-$ ,  $Cl^-$ , and water across Fluc-Bpe while being open and filled with water molecules.  $z = 0$  corresponds to the COM of the transmembrane residues in Fluc-Bpe. The position of the F82 and F85 binding sites are highlighted. (A) Cycled and bootstrapped PMFs. (B) Non-cycled PMFs

TM3b due to the large amount of water molecules solvating the ions. In contrast, a RMSD value of  $\sim 0.1$  nm indicates that the channel structure barely changes along the simulations in which water is bound to Fluc-Bpe as compared with the crystal. Taken all together, these results indicate that the structure of the proposed permeation pathway in Fluc-Bpe is significantly different to the crystal in simulations with both polarizable and non-polarizable force-fields upon restraining the fluoride ions inside. Besides, the channel structure remains similar to the crystal when water molecules are at the binding sites. Overall, these results suggest that the functional conductive-state in Fluc may be an open and partially hydrated channel.

## 9.2 PMFS OF SOLUTE PERMEATION ACROSS AN OPEN FLUC-BPE

To investigate the permeability of a Fluc-Bpe channel that is *open* and fully hydrated, we computed the PMFs for the permeation of  $F^-$ ,  $Cl^-$ , and water with US simulations (see Section 6.5.2) (see Figure 9.3). The PMF for the permeation of water exhibits a free energy barrier  $\Delta G^\ddagger$  of  $\sim 7.5$   $\text{kJ mol}^{-1}$  close to the F82 binding site, which is compatible with a fully hydrated pore and a highly efficient water channel such as the ones previously analyzed in this thesis, i. e. UT-B and AQP1. At the F85 binding site, the water PMF is  $\sim 5$   $\text{kJ mol}^{-1}$  lower than at F82. This is consistent with the free energy calculations seen in the previous chapter, which resulted in a absolute binding free energy  $\Delta G_{\text{bind}}$



at F85 of  $\sim 10 \text{ kJ mol}^{-1}$  lower than at F82 (see Section 8.3). The PMFs for fluoride and chloride permeation showed in Figure 9.3A present free energy barriers  $\Delta G^\ddagger$  of  $\sim 18 \text{ kJ mol}^{-1}$ , which is compatible with the permeation of these ionic species. These results indicate that the Fluc-Bpe open structure may be the conductive state of the channel. However, some uncertainties regarding the permeation and selectivity mechanisms remain. For instance, the energy barrier in the PMF of  $\text{F}^-$  is slightly higher than in the PMF of chloride, indicating that this *open* Fluc-Bpe is not selective to  $\text{F}^-$  over  $\text{Cl}^-$ , in contrast with experimental evidence [74]. At the F85 binding site, the PMFs for  $\text{F}^-$  and  $\text{Cl}^-$  exhibit a steep increase, with energy values similar to those of the PMF for water. These values reflect the permeation of an anion that is partially or fully hydrated, and clearly diverge from the relative binding free energies  $\Delta\Delta G_{\text{bind}}$  between water and fluoride, in which water is favored by at least a  $40 \text{ kJ mol}^{-1}$  difference (see Figure 8.2).

The PMFs in Figure 9.3A have been calculated using a cyclic version of WHAM, which renders a periodic PMF  $G(z)$  within the interval  $[z_{\text{min}}, z_{\text{max}}]$ , as expected for the permeation in case of equal chemical potential in the two water compartments [181]. The PMFs in Figure 9.3B are not cycled, thereby showing offsets of  $\sim 15$  to  $20 \text{ kJ mol}^{-1}$  for fluoride and chloride, respectively, between each side of the membrane, demonstrating that the PMFs are not converged. In general, these results suggest that i) the crystal structure of Fluc-Bpe is most likely a closed pore, and that ii) we derived a conductive state of Fluc as confirmed by a low free energy barrier for  $\text{F}^-$  permeation. However, some key challenges remain. Most importantly, the mechanism of selectivity of  $\text{F}^-$  over  $\text{Cl}^-$  is not yet understood. We speculate that force-fields that accurately model the anion-protein interactions will be required to achieve this goal.



Part IV

DISCUSSION AND CONCLUSIONS



---

## DISCUSSION

---

### 10.1 IS 3D-RISM MORE SUITABLE THAN MD TO CALCULATE PMFS OF SOLUTE PERMEATION?

We have presented a systematic comparison between two different methods for the calculation of PMFs of solute permeation across membrane channels, namely, 3D-RISM and MD simulations via Umbrella Sampling (US).

For water, we found significant agreement between the PMFs calculated with 3D-RISM and MD simulations, supporting many other studies in which 3D-RISM has been used to calculate water density distributions in biological systems [252–257]. The small discrepancies between the water PMFs from MD and 3D-RISM may be ascribed to i) the lack of protein structural fluctuations in 3D-RISM calculations; ii) the different water models used in MD simulations and 3D-RISM [198]; iii) the sampling limitations inherent to MD; and iv) the orientational averaging carried out in RISM calculations.

In contrast to our findings for water, for the rest of the permeating solutes we found significant deviations between the PMFs obtained from MD and 3D-RISM. The PMFs calculated from MD simulations are determined by the polarity, size, and shape of the solute, responding to different solute-protein interactions. For instance, the PMFs exhibit lower profiles in the apolar regions of the pore in which certain solutes establish partially apolar contacts with the channel (e.g. O<sub>2</sub>, methanol, and urea), thereby replacing unfavorable water-protein interactions. Many other authors have calculated similar PMFs from MD simulations, which are consistent with our results [62–64, 244, 258]. By contrast, the PMFs computed from 3D-RISM did not reveal any of the features arising from the chemical specificities of the solutes. The different PMFs calculated from 3D-RISM were too similar. The only variations in the profiles seem to originate from solute size but not from solute polarity, thereby missing any modulation due to hydrophobic effects. The lack of hydrophobic contributions to 3D-RISM has been noticed before, and thus several corrections that allow the calculation of more accurate solvent Radial Distribution Functions (RDFs) around a hydrophobic solute, such as the ones indicated by Cao *et al.*, might improve the PMFs calculated from 3D-RISM [259].

One of the most striking deficiencies exhibited by the PMFs computed from  $3D$ -RISM for urea permeation across UT-B is the lack of urea binding sites (Figures 7.1A, 7.4A/B, 7.5E, 7.6C, 7.7E, 7.8A, and 7.9). Apparently,  $3D$ -RISM fails to capture the different roles that urea may play at its binding sites, i. e. as an h-bond donor and acceptor. Urea is a planar molecule whose edges are polar whilst the faces are apolar. These particular features are leveraged in the urea binding site at  $S_0$  shown in Figure 10.1A/B, in which urea establishes an h-bond with residue T334 as acceptor, and with residues I228 and Q227 as donor. Simultaneously, the planar and less hydrophylic faces may interact favorably with the apolar residues F283, L123, L127, and L287. These residues form a highly optimized binding site, in which urea is coordinated by a combination of polar and apolar contacts leading to an energetically stabilized pose, as reflected by the minimum of  $\sim -5 \text{ kJ mol}^{-1}$  in the PMF calculated from MD (Figure 7.1A, green curve). For comparison, Figure 10.1C/D shows the binding site at  $S_0$  as interpreted by  $3D$ -RISM, with the density distribution of the urea carbon atom depicted in turquoise. Notably,  $3D$ -RISM suggests reduced density at the binding site, incompatible with the binding of urea. These results suggest that, given the relevance of the orientational dependence of certain solutes to establish stable solute-protein interactions, the final 3-dimensional density distributions of solvent molecules across protein channels may be critically affected by the loss of information owing to the orientational averaging performed by DRISM [190, 193].

We further tested the effect of several parameters relevant to DRISM and  $3D$ -RISM calculations, such as water model, solvent site, grid spacing, solute concentration, and error tolerances. None of these had a significant effect on the resulting PMFs. However, changes in the solvent site or the grid spacing may lead to larger variations and should, therefore, be tested with care (Figures 7.7, and 7.8). Additionally, we tested different closure relations, such as HNC and PSE- $n$ , as an alternative to the commonly used KH closure, for both DRISM and  $3D$ -RISM calculations. Unfortunately, converging the equations with HNC or PSE- $n$  closures is more difficult as compared with the KH closure [193, 260]. Besides, when the HNC and PSE- $n$  solutions converge, the resulting PMFs resembled the PMFs obtained with the KH closure (Tables 6.1, and 6.2; Figures 7.5, and 7.6).

Furthermore, we tried a different approach, which consisted in the computation of PMFs for the permeation of non-water solutes using the hydration free energies calculated from  $3D$ -RISM. In this approach, the protein *and* the permeating molecule were defined as the “solute” according to the  $3D$ -RISM conventions, and water was defined as the only “solvent”. Then, the hydration free energy of the solute —protein *and* permeating molecule— is computed with  $3D$ -RISM as a function of the permeating molecule position across the chan-

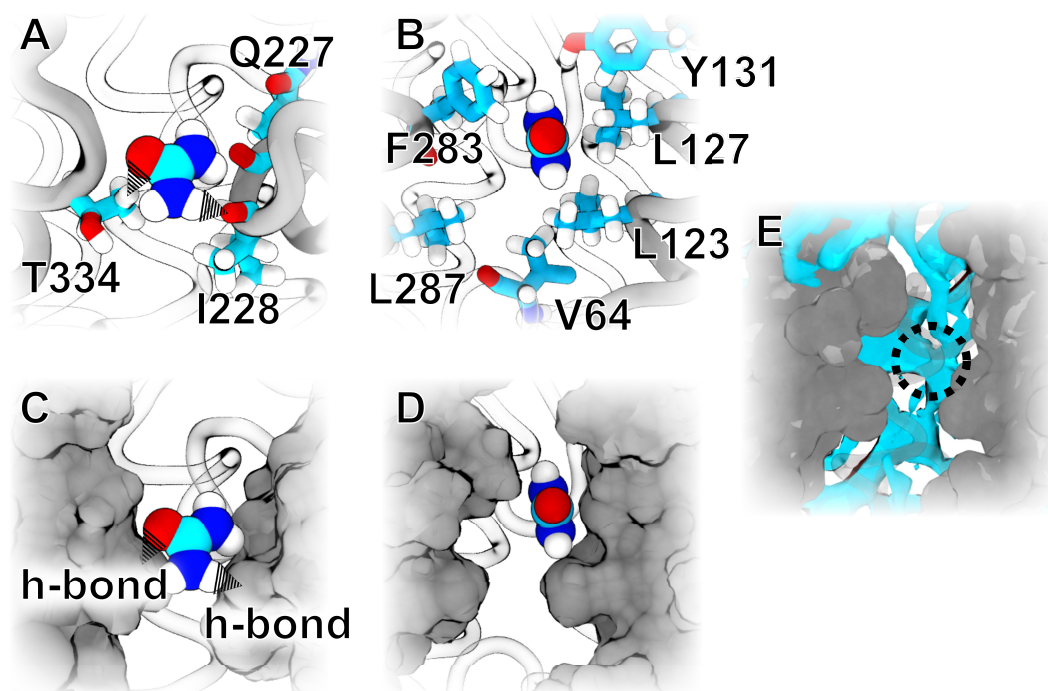


Figure 10.1: Urea binding site in UT-B  $S_0$  region. The protein secondary structure and volume is shown in silver, urea as spheres, and some side chains as thick sticks. The color coding for the atoms is as follows: carbon (turquoise), oxygen (red), nitrogen (blue), and hydrogen (white). The snapshots were taken from US MD simulations. (A/C) Hydrogen bonds formed between the urea oxygen and T334 hydrogen, as well as between a urea hydrogen and I228 oxygen are depicted in black dashed line. (B/D) 90 deg rotation of A, illustrating favorable apolar contacts between urea and residues F283, L287, L123, and L127. V64 and Y131 are also depicted filling the space. (E) 3D density of the urea carbon from  $3D-RISM$ . The turquoise surface indicates  $g(r) > 2$ , and the void space in the density distribution function is highlighted with a dotted circle.

nel  $z$ , being performed over many conformations of the permeating molecule and finally averaged. These calculations, however, poorly converged, even when the protein coordinates were frozen, indicating that 3D-RISM hydration free energy based calculations are not an alternative to 3D-RISM calculations with solvent mixtures.

3D-RISM was previously used to calculate PMFs of solute permeation across AQP<sub>1</sub>. In 2010, Phongphanphanee *et al.* computed the PMFs for the permeation of water, Ne, CO<sub>2</sub>, NO, NH<sub>3</sub>, urea, and glycerol across AQP<sub>1</sub> [99]. However, our PMFs computed from 3D-RISM differ from their PMFs, deserving more explanations. First, the free energy barriers of 10 to 12 kJ mol<sup>-1</sup> present in our PMFs of water permeation cannot be observed in the profiles computed by Phongphanphanee *et al.*, which are, to a large extent, flat. The free energy barriers of at least 10 kJ mol<sup>-1</sup> do not reflect a channel empty of water, instead, they account for the entropic penalty due to the narrowing of the pore as compared with the bulk region [261]. Indeed, our PMFs of water permeation across UT-B and AQP<sub>1</sub> are the result of water-filled channels (Table 7.1). Before translating the 1D density profile  $g(z)$  into a PMF (Section 6.2.1), Phongphanphanee *et al.* divided  $g(z)$  by the cross section area of the pore (Equation 4 in ref. [262]), yielding a flat profile. This corresponds to the solute density *only* inside the channel, instead of to the solute density along the reaction coordinate  $z$ , leading to profiles that are not PMFs as such and, therefore, cannot be compared with the PMFs presented in this study.

## 10.2 THE NATURE OF THE CRYSTALLOGRAPHIC ELECTRON DENSITIES IN FLUC-BPE

Many structural and functional properties of proteins depend on water-protein and ion-protein interactions. The latter play an essential part in many enzymatic reactions, shaping the free-energy landscape of chemical reactions, and by stabilizing structural motifs. Many of these relevant ions and water molecules are detected in crystallographic electron densities, the proper assignment of which is pivotal for deriving functional mechanisms from structural data. However, the correct assignment of densities is not straightforward and often requires a considerable amount of human expertise [48, 263–272]. In this thesis we used a combination of MD simulations and free energy calculations to assign the five disconnected electron densities in the crystal structure of the bacterial fluoride channel Fluc-Bpe.

Overall, our results suggest that the four disconnected electron densities, provisionally assigned to F<sup>-</sup> in the crystal structure by Stockbridge *et al.*, are most likely ordered water molecules. Regarding the fifth density, which is at the center of the channel (TM3), the relative binding free energy calculations yielded inconclusive results. Consequently, short free MD simulations were used to evaluate the struc-



tural stability of the TM3 binding site, corroborating the assignment of  $\text{Na}^+$ , as indicated by the authors of ref. [85].

The limited precision of ion-protein interactions in additive force-fields proved to be a source of uncertainty, which makes the absolute and relative binding free energy calculation results highly dependent on the applied force-fields. For example, the interactions between a cation and carboxylate along distance in vacuum, calculated with both quantum-chemical calculations and older versions of common force-fields, may differ by tens of  $\text{kJ mol}^{-1}$ , which triggered the development of several force-field fixes in the last years [232–234]. Here, we computed the relaxed Potential Energy Surface (PES) along fluoride-benzene and chloride-benzene distances in vacuum with quantum-chemical calculations and compared them with the profiles derived from a non-polarizable force-field (Amber ff99SB\*-ILDN) and a polarizable force-field (CHARMM-Drude) (see Figure 8.6). In both cases (fluoride-benzene and chloride-benzene), the results calculated with the Amber force-field in combination with the ion parameters from Joung & Cheatham III underestimated the anion-benzene interactions by tens of kilojoules per mol, whilst the profiles derived from CHARMM-Drude were fairly consistent with the PESs from QM calculations. These results indicate the appropriateness of using polarizable force-fields when ion-protein interactions come into play. However, since the relative binding free energy differences  $\Delta\Delta G_{\text{bind}}$  between  $\text{F}^-$  and water at F82 and F85 are larger than the inaccuracies derived from the non-polarizable force-field, and given that fluoride is equally unstable at the binding sites with CHARMM-Drude as it is with Amber99SB, the qualitative conclusions regarding the nature of the electron densities seem to be unaffected by the force-field defects.

One could argue about the convenience of simulating in a physiologically realistic environment instead of the crystallographic unit cell. On one hand, the simulations with the crystallographic unit cell would be carried out in conditions more similar to experiment and, in addition, such simulations would allow high-throughput free energy calculations. On the other hand, a system including the lipid bilayer, solvated with explicit water molecules and counter-ions, allowed us to neutralize the overall charge while transforming a water molecule into fluoride in bulk, and fluoride into a water molecule in the protein, during the alchemical transformation simulations. With a non-neutral box, the PME method for computing the electrostatic interactions would have built up a so-called background charge density, whose interactions with the system ions lead to artifacts [273]. However, our simulation set-up, which includes the lipidic membrane and explicit solvent molecules, is not suitable for high-throughput free energy calculations. It is important to stress, however, that alchemical transformations in the unit cell are not possible under neutral overall charge conditions, unless enough volume of bulk water is provided

to add counter-charges. Therefore, to perform such calculations in the unit cell, it would be desirable to apply analytical corrections similar to the ones already developed for hydration free energy calculations of ions in a uniform solvent, or even for simple heterogeneous systems, such as hydrophobic slaps or spheres in water, which alleviate the effects due to the background charge, the finite size of the simulation system, or the finite volume of the ion [96, 273–281]. Although such corrections are not suitable for more complex heterogeneous systems such as protein crystals, we hope that approximations might be established based on the above-mentioned studies.

### 10.3 THE PERMEATION MECHANISM IN FLUC-BPE

We used a combination of free MD simulations and US simulations to test the suggested permeation mechanism in Fluc-Bpe by Stockbridge *et al.* The authors proposed that fluoride permeation takes place through a fully dehydrated pore, in which the contacts between the permeating ion and the residues that form the polar-track of the protein, including the quadrupolar interactions with F82 and F85, play a major role [85, 98].

Our results suggest that Fluc-Bpe undergoes a pore *opening* transition—regardless of the applied force-field— if fluoride ions are restrained at F82 and F85. This transition takes place after water molecules from bulk solvent enter the channel to hydrate the ions. In contrast, when water molecules are bound at F82 and F85, no additional water molecules enter Fluc, and its structure remains similar to the crystal. Many ion channels need a transition from a close- to an open-state to facilitate the ionic flux [47, 282–285]. In the case of Fluc channels, single channel current experiments indicate that  $F^-$ -flux is independent from gating events [74]. The absence of gating events does not mean that there are not open and closed states, it rather means that the functional state in Fluc is mainly open or conductive. In its crystallographic form, Fluc is bound to pore-blocking monobodies, which may affect the functional state of the channel, thereby leaving Fluc-Bpe in a close- or non-conductive state. However, recent analyses of pore-blocking monobody binding to Fluc-Bpe rule out global and local allosteric mechanisms that would change the Fluc structure significantly [87, 88, 286]. But in the study by McIlwain *et al.*, the Fluc channel blocked by one monobody has a conductivity up to 10-15% of unblocked Fluc-Bpe. Hence, we cannot ignore the possibility of an alternative fully conductive *open*-state. In our simulations, we observed *opening* transitions when the  $F^-$  ions were restrained at F82 and F85. It is important to highlight that these results were replicated in the simulations with the polarizable CHARMM-Drude force-field indicating that, even when ion-protein interactions are accurately modelled, Fluc-Bpe cannot conduct dehydrated ions. Overall, the qualita-

tive agreement between the results yielded by non-polarizable and polarizable force-fields suggests that reaching the conductive state of Fluc implies structural changes that lead to partially hydrated pores.

Accordingly, we calculated the PMFs for solute permeation with this *open* structure (Figure 9.3). The PMF for water suggests that water would permeate Fluc-Bpe as efficiently as other well-known water channels such as UT-B or AQP1. For  $F^-$  and  $Cl^-$ , the PMFs exhibit a free energy barrier compatible with the permeation of either ion. Hence, the PMFs do not reproduce the experimentally established selectivity for  $F^-$ . This could be explained by the balance between ion-water and ion-protein interactions. In the case of non-polarizable additive force-fields, this balance clearly favours the ion-water interactions. Hence, key ion-protein contacts might be absent in our simulations, in which the ions are constantly solvated along the permeation pathway. The PMFs for ion permeation did not fully converge, leaving room for improvement. On one hand, it is not surprising that water molecules may permeate Fluc, as similar mechanisms have been observed in other ion-channels such as  $K^+$ -channels and  $Na^+$  channels [287–290]. On the other hand, the PMFs do not reflect a clear selectivity for  $F^-$  over  $Cl^-$ , raising doubts about the robustness of the conclusions on ion permeation across Fluc. We speculate that, given that the fluoride-Phe / chloride-Phe interactions as modelled by additive force-fields are underestimated, the usage of the more accurate CHARMM-Drude force-field will be critical for analyzing the function of Fluc-Bpe in particular, and ion-channels in general.

The permeation mechanism proposed by Stockbridge *et al.* and by Last *et al.* implies that a dehydrated fluoride ion would be permeated by a Fluc channel that is structurally similar to the crystal. Then, fluoride would establish key contacts with certain residues that form a polar-track across the pore, in which the conserved phenylalanine residues would have a preferent role. This mechanism contrasts with the results of our simulations. In the simulations fluoride rapidly escapes Fluc when in direct contact with the phenylalanines. In addition, when fluoride is restrained inside the channel, it drags water into the pore leading to a conductive open state. These results can be reproduced with the polarizable CHARMM-Drude force-field, by which the anion-phenylalanine interactions are more accurately modelled as compared with non-polarizable additive force-fields. A recent study has measured the magnitude of the anion-quadrupole interactions in protein residues, which range from  $\sim -60$  to  $0 \text{ kJ mol}^{-1}$ , highlighting the relevance of such interactions [291]. Hence, we agree with Stockbridge *et al.* in the critical role fluoride-phenylalanine contacts may play for permeation. We add that the selectivity mechanism might be rationalized by the potential energy difference between fluoride-benzene and chloride-benzene interactions, as revealed by the PES scans. Thanks to this difference, which was of

$\sim 20 \text{ kJ mol}^{-1}$ , the more favorable fluoride-phenylalanine interactions, as compared with the chloride-phenylalanine ones, may help Fluc channels to facilitate the selective flux of  $\text{F}^-$  over  $\text{Cl}^-$ .

Our insight into the functioning of Fluc channels can be summarized as follows: i) The conductive state of Fluc channels entails a pore opening transition leading to an open structure that is different from the structure resolved by crystallography; ii) water molecules populate the open pores in Fluc, partially solvating the permeating fluoride ions; iii) the selective permeation of fluoride in Fluc channels may arise from favorable  $\text{F}^-$ -Phe edgewise contacts which, being stronger than the  $\text{Cl}^-$ -Phe interactions, would efficiently discriminate fluoride from chloride. Several questions about the permeation and selectivity mechanisms in Fluc channels remain open nevertheless:

- *Is the open state a plausible state in conductive Fluc?* Since we did not observe any opening transition unless the fluoride ions were restrained inside, we cannot rule out the possibility that the open state is an artifact caused by the presence of  $\text{F}^-$  inside Fluc. Therefore, to confirm the plausibility of the conductive *open* state on Fluc, it would be desirable to reproduce the opening transition in the absence of restrained fluoride ions. Probably this would require long MD simulations in the range of, at least, tens of microseconds. In addition, Nuclear Magnetic Resonance (NMR) spectroscopy might help to characterize the structural properties of Fluc in native-like conditions, thereby confirming the open conductive state suggested in this thesis. NMR has been already used for membrane proteins [292, 293]. However, the classical protein size limit of  $\sim 30 \text{ kDa}$  in NMR studies might be too tight even for the relatively small bacterial Fluc channels ( $\sim 32 \text{ kDa}$ ). A suitable alternative, then, may be solid-state NMR, which does not suffer from the classical size restrictions and has been already used in structural studies with liposomes [294, 295].
- *Is there an alternative permeation pathway?* The permeation pathway in Fluc channels has been suggested according to i) the position of the four electron densities that were assigned to fluoride in the crystal, and ii) to the fluoride flux disruption upon mutating the conserved Phe residues. It is clear now that the electron densities correspond to ordered water molecules, so the first clause can be ignored. Hence, the question remaining is what different permeation pathways, dependent on the conserved phenylalanines, are still possible. To tackle this question we suggest to perform fluoride steered MD simulations across Fluc. In theory, a sufficiently slow pulling should allow the ion being pulled to find the less restrictive path across the channel. Such approach has been already used to understand the molecular mechanism of ion permeation in different systems. Steered

MD simulations can be used as well to calculate PMFs of ion permeation, which would give a more complete picture of the permeating process [296–298]. Since ion-protein interactions are crucial in ion permeation processes, the use of polarizable force-fields will be relevant to account for the proper magnitude of ion-protein energetic contributions.

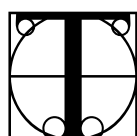
- *What is the role of water in the channel?* If the conductive state of Fluc is not the open state, but a close state similar to the crystal, then the four ordered water molecules observed in the crystal could play an important structural role. For instance they might fix the position of the conserved Phe residues so that the permeating fluoride ion establishes properly oriented edgewise contacts. If, instead, the conductive state of Fluc is similar to the hydrated open state observed in the simulations, then the main role of water molecules might be to lubricate the pore. Thus water would partially hydrate the permeating fluoride ions which, otherwise, would have to overcome a dehydration penalty of  $\sim 500 \text{ kJ mol}^{-1}$ . Without a detailed structural characterization of the conductive state of Fluc, this question promises to be quite difficult to answer. Even in this scenario, the role of water in ion channels still sparks lively debates that are far from being closed. This is the case with potassium channels, in which it is not clear whether water molecules participate in the permeation mechanism [48, 267, 268, 285, 299].
- *Which is the selectivity mechanism?* Regardless of the open or closed form of its conductive state, the way Fluc selectively permeates fluoride and excludes chloride remains a mystery. On one hand, the ionic radius of  $\text{F}^-$  (1.2 Å) is slightly smaller than that of  $\text{Cl}^-$  (1.7 Å). This difference could be exploited by Fluc channels to discriminate between the two species. Similar molecular filtering has been proposed for other channels such as  $\text{Ca}^{2+}$  release-activated  $\text{Ca}^{2+}$  (CRAC) channels or 5-HT<sub>3A</sub> receptors. The structural characteristics of such channels determine the size of the largest permeating ion that can diffuse through them [300, 301]. On the other hand, as suggested by the PES calculations, the difference of  $\sim 20 \text{ kJ mol}^{-1}$  in the interaction energy between  $\text{F}^-$ -Phe and  $\text{Cl}^-$ -Phe might be enough to allow the efficient and selective fluoride flux. We speculate that the selectivity in Fluc channels might stem from one of the discussed mechanisms or the combination of both. To answer this, it would be required the computation of PMFs for ion permeation with different ionic species using polarizable force-fields.



---

 CONCLUSION
 

---



THIS THESIS has clarified an open question about the suitability of  $3D$ -RISM to calculate PMFs for solute permeation across protein channels. It has proposed a new method to ascertain the nature of isolated electron densities in protein crystals, and it has developed a solid hypothesis about the permeation mechanism and selectivity in Fluc channels.

We found in  $3D$ -RISM a computationally cheap and reasonably accurate method to compute PMFs for water permeation. On one hand, free energy differences of  $\sim 2 \text{ kJ mol}^{-1}$  between the UT-B PMFs yielded a 50% lower water occupancy in  $3D$ -RISM as compared with MD. On the other hand, the agreement between  $3D$ -RISM and MD for the AQP1 PMFs is nearly quantitative. Therefore,  $3D$ -RISM stands as an efficient alternative to MD simulations for water-density predictions. In contrast, the treatment of non-water solutes in  $3D$ -RISM calculations led to PMFs which significantly diverged from those computed with MD, even when different closures, solvent-sites, or the structural fluctuations of the protein were considered in  $3D$ -RISM. The origin of such divergence might be the result of i) the intrinsic limitations imposed by the orientational averaging carried out in RISM, and ii) the incorrect treatment of hydrophobic interactions in  $3D$ -RISM.

We further proposed and demonstrated a new method based on MD simulations and free energy calculations for assigning crystallographic electron densities. We found that i) *relative* binding free energies  $\Delta\Delta G_{\text{bind}}$  between two different molecular species may determine which species binds with higher affinity to a certain binding site. However, when the balance between ion-protein and water-protein interactions is involved in  $\Delta\Delta G_{\text{bind}}$  calculations, current additive force-fields, in which ion-protein interactions are poorly modelled, do not provide accurate results. Therefore, conclusions may only be drawn when the  $\Delta\Delta G_{\text{bind}}$  values clearly deviate from zero, as they did for the  $F^-$ /water alchemical transformations. ii)  $\Delta\Delta G_{\text{bind}}$  clarifies whether the binding of one species is more plausible over another. However,  $\Delta\Delta G_{\text{bind}}$  does not confirm whether one species binds *per se*. The binding can only be confirmed upon the calculation of *absolute* binding free energies  $\Delta G_{\text{bind}}$ , which reveals whether a certain species *actually* binds to a particular binding site. Binding is only plausible when

$\Delta G_{\text{bind}} \lesssim 0$ , which is needed for a molecule to be observed in the crystal. Furthermore, the fraction of binding sites occupied by that molecule also depends on its bulk concentration. iii) Short, free MD simulations may be useful for analyzing the structural stability of the crystal. Fast and reproducible deviations from the crystal structure may indicate wrong assignment of electron densities. In our case, the recurrent instability of fluoride ions at the F82 and F85 binding sites, regardless of the applied force-field, and the systematic drift in the TM3 site upon binding of water, helped to determine the nature of the densities in those sites. It is important to stress, however, that the opposite is not true, i. e. the structural stability in short, free MD simulations does not confirm an assignment, since many structural deviations may be random and occur beyond the available time scale. Hence, solid conclusions may only arise from multiple independent simulations that corroborate comparative stability analyses.

Finally, we elaborated a hypothesis about the permeation mechanism and selectivity in Fluc channels. We found that the presence of fluoride ions inside the pores transforms Fluc into an open, hydrated channel, which is structurally different to the crystal. We therefore conclude that Fluc channels undergo an opening transition to reach the conductive functional state. According to the calculated PMFs for solute permeation, the open-state is permeable to water, fluoride and chloride. Thus, the experimentally observed selectivity for fluoride was absent in our model. In addition, the PES scans revealed a difference of  $\sim 20 \text{ kJ mol}^{-1}$  between  $\text{F}^-$ -Phe and  $\text{Cl}^-$ -Phe contacts. Hence, we propose that the specificity for fluoride in Fluc channels stems from favorable  $\text{F}^-$ -Phe edgewise-interactions.



Part V

APPENDIX



# A

---

## ENTROPY IN A FLAT-BOTTOMED CYLINDRICAL POTENTIAL

---

During US simulations of solute permeation across UT-B (see Section 6.1 for details), the solutes were restrained to a cylinder via a cylindrical flat-bottomed potential  $V_{\text{cyl}}(r) = k_c(r - r_c)^2/2 \cdot H(r - r_c)$ . Thus, when solutes are in bulk, they may diffuse laterally along an area defined by the cross section area  $A_{\text{cyl}}$  of the cylinder. Therefore, the corresponding PMF relates to a density of one channel per cross section area of the cylinder. In other words, the cross section area of the cylinder might be regarded as the bulk reference area of the PMF. Hence, to translate the PMF into a new reference area  $A_{\text{ref}}$ , a correction for the change in entropy of the bulk upon changing the reference area from  $A_{\text{cyl}}$  to  $A_{\text{ref}}$  has to be applied. That change in entropy is defined as

$$\Delta S = k_B \ln \left( \frac{A_{\text{ref}}}{A_{\text{cyl}}} \right). \quad (\text{A.1})$$

For a “hard-core” cylinder, the cylinder area would be simply  $A_{\text{cyl}}^{\text{hc}} = \pi r_c^2$ . However, for a “soft” cylinder, the entropy of the solute density defined by the flat-bottomed potential has to be calculated. The solute density is

$$\rho_{\text{cyl}}(r) = N^{-1} \begin{cases} 1 & \text{if } r \leq r_c, \\ \exp\left(-\frac{(r-r_c)^2}{2\sigma_c^2}\right) & \text{if } r > r_c, \end{cases} \quad (\text{A.2})$$

where  $\sigma_c = (k_B T/k_c)^{1/2}$  is the width of the Gaussian-shaped decay of the density in the quadratic region of the flat-bottomed potential. The normalization constant is

$$N = \pi \left( r_c^2 + 2\sigma_c^2 + \sqrt{2\pi} r_c \sigma_c \right), \quad (\text{A.3})$$

such that  $\rho_{\text{cyl}}(r)$  is normalized,  $\int_0^\infty 2\pi r \rho_{\text{cyl}}(r) dr = 1$ . The Shannon entropy of the distribution is

$$\begin{aligned} S_{\text{cyl}}[\rho_{\text{cyl}}] &= -k_B \int_0^\infty 2\pi r \rho_{\text{cyl}}(r) \ln \rho_{\text{cyl}}(r) dr & (\text{A.4}) \\ &= \frac{k_B \pi r_c^2}{N} \ln N \\ &\quad + \frac{2k_B \pi \sigma_c}{N} \left[ \sigma_c (1 + \ln N) + r_c \sqrt{\frac{\pi}{2}} \left( \frac{1}{2} + \ln N \right) \right]. & (\text{A.5}) \end{aligned}$$

The Shannon entropy of a uniform distribution over the new reference area is

$$S_{\text{ref}} = k_B \ln A_{\text{ref}}. \quad (\text{A.6})$$

Thus, the corrected free energy in bulk is defined as

$$\Delta G_{\text{bulk}} = -T(S_{\text{ref}} - S_{\text{cyl}}), \quad (\text{A.7})$$

where  $T$  is the absolute temperature. Since the free energy of the bulk state is often defined to zero, it is convenient to apply the correction to the PMF in the inside region of the channel by  $-\Delta G_{\text{bulk}}$ . Otherwise the bulk region should be corrected by  $\Delta G_{\text{bulk}}$ .

Taking numerical example, let  $r_c = 0.7$  nm,  $k_c = 500$  kJ mol<sup>-1</sup> nm<sup>-2</sup>,  $T = 300$  K, and  $A_{\text{ref}} = 10.3$  nm<sup>2</sup>, which is the cross section area of an AQP1 monomer [172]. Then, we calculate  $\Delta G_{\text{bulk}} = 3.851$  kJ mol<sup>-1</sup>. Approximating the cylinder area as  $A_{\text{cyl}} = \pi(r_c + 2\sigma_c)^2$  yields a correction of 3.824 kJ mol<sup>-1</sup>, which is consistent with the analytic result, as suggested before [212]. On the contrary, approximating the cylinder area as  $A_{\text{cyl}} = \pi r_c^2$ , and thus neglecting the density outside of the flat region of the flat-bottomed potential, yields an inaccurate correction of 4.74 kJ mol<sup>-1</sup>.

# B

---

## TIME TRACES OF $F^-$ AND WATER POSITIONS INSIDE FLUC-BPE

---

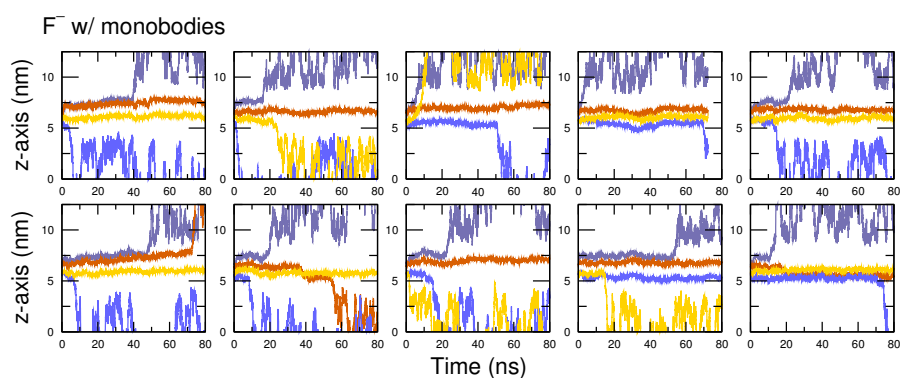


Figure B.1:  $F^-$  positions in Fluc-Bpe blocked by L2 monobodies. Orange and blue curves represent the molecules starting at the F82 and F85 sites, respectively. Different shades of orange and blue indicate starting positions in the two homodimers of Fluc-Bpe.

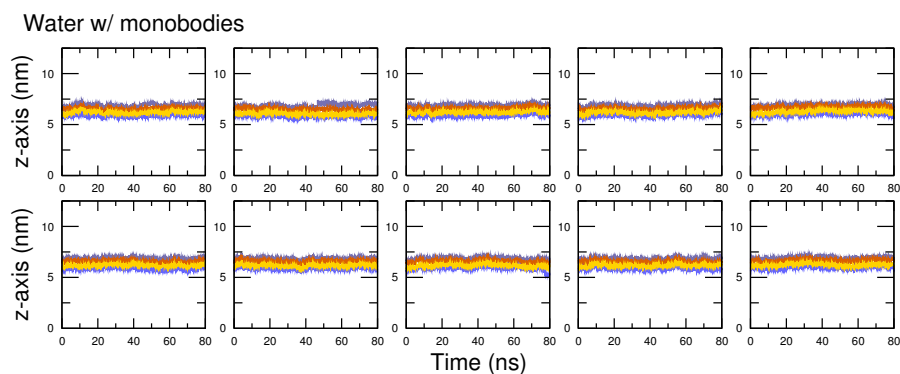


Figure B.2: Water molecule positions in Fluc-Bpe blocked by L2 monobodies. Orange and blue curves represent the molecules starting at the F82 and F85 sites, respectively. Different shades of orange and blue indicate starting positions in the two homodimers of Fluc-Bpe.

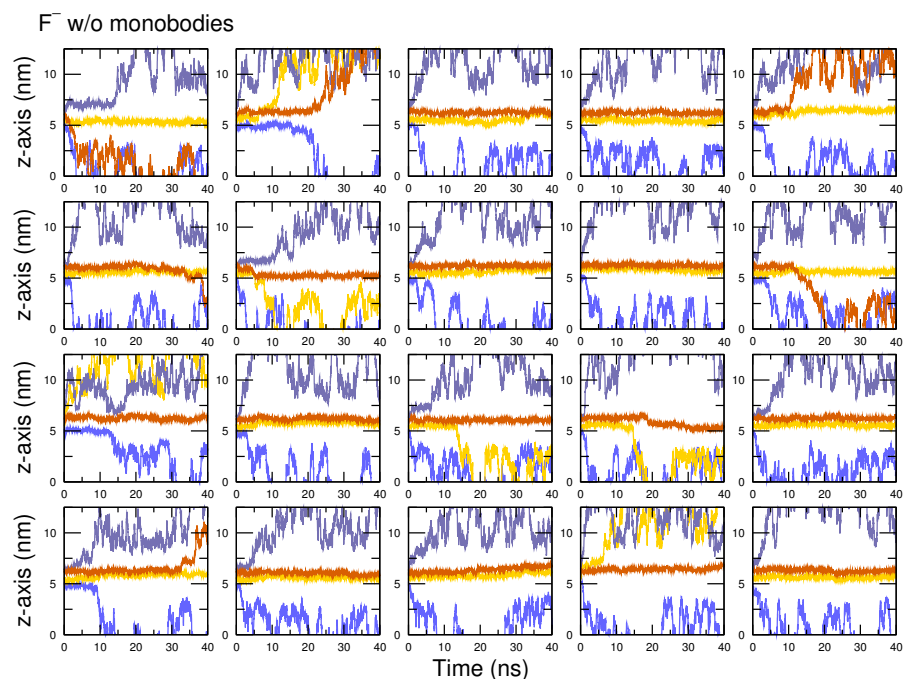


Figure B.3:  $F^-$  positions in Fluc-Bpe free of monobodies. Orange and blue curves represent the molecules starting at the F82 and F85 sites, respectively. Different shades of orange and blue indicate starting positions in the two homodimers of Fluc.

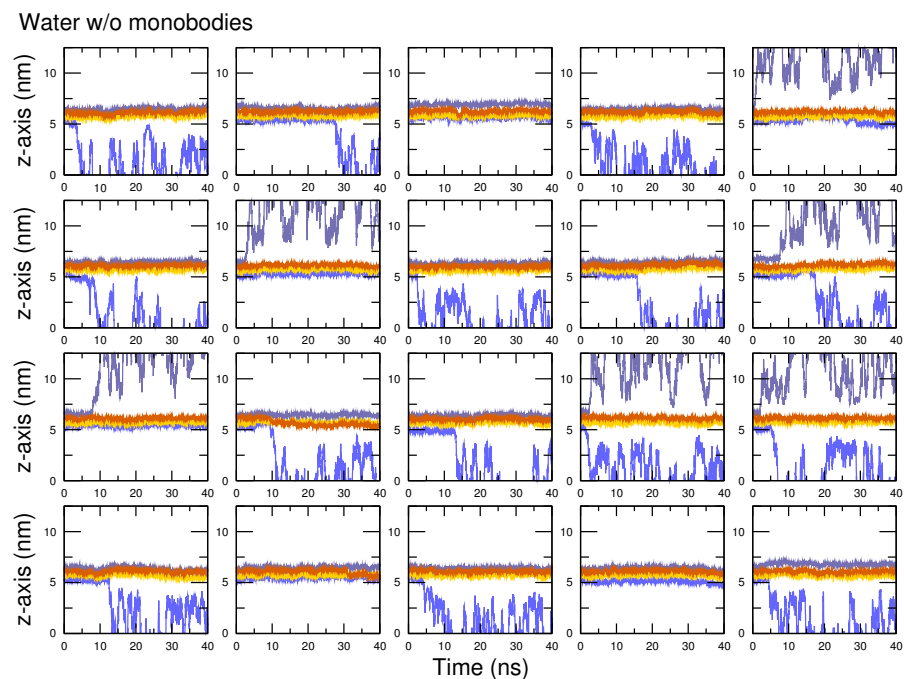


Figure B.4: Water molecule positions in Fluc-Bpe free of monobodies. Orange and blue curves represent the molecules starting at the F82 and F85 sites, respectively. Different shades of orange and blue indicate starting positions in the two homodimers of Fluc-Bpe.

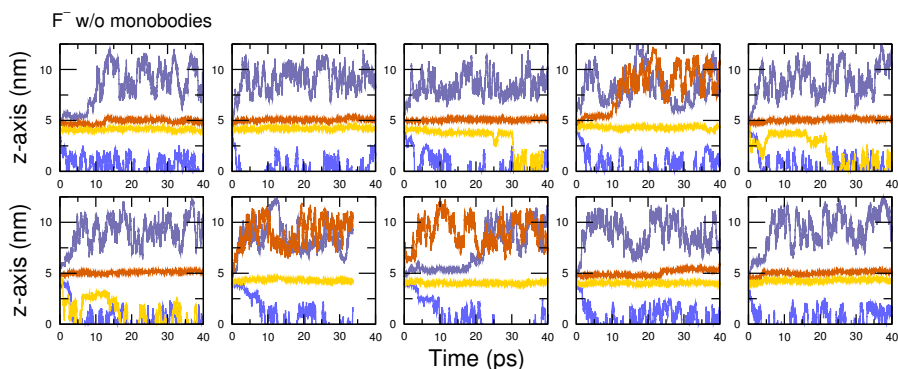


Figure B.5:  $F^-$  positions in Fluc-Bpe free of monobodies, simulated with the CHARMM36 force-field for the protein and Joung/Cheatham parameters for  $F^-$ . Orange and blue curves represent the molecules starting at the F82 and F85 sites, respectively. Different shades of orange and blue indicate starting positions in the two homodimers of Fluc-Bpe.

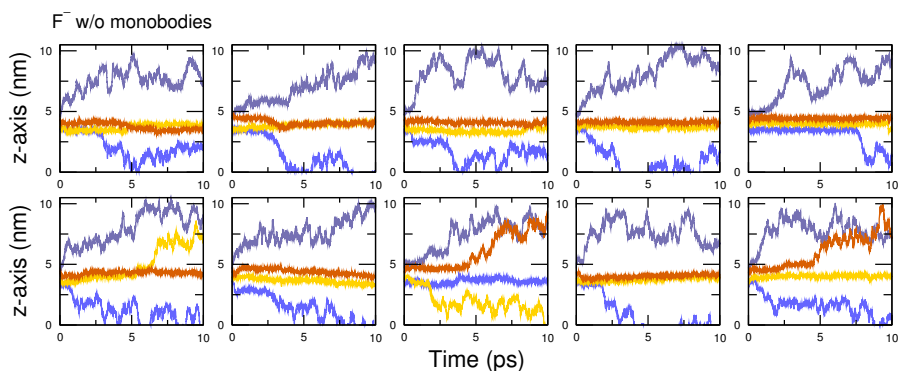


Figure B.6:  $F^-$  positions in Fluc-Bpe free of monobodies, simulated with the CHARMM-Drude force-field for the protein and the polarizable fluoride model by Lamoureux & Roux. Orange and blue curves represent the molecules starting at the F82 and F85 sites, respectively. Different shades of orange and blue indicate starting positions in the two homodimers of Fluc-Bpe.





# C

## TIME TRACES OF THE AVERAGE NUMBER OF WATER MOLECULES INSIDE FLUC-BPE

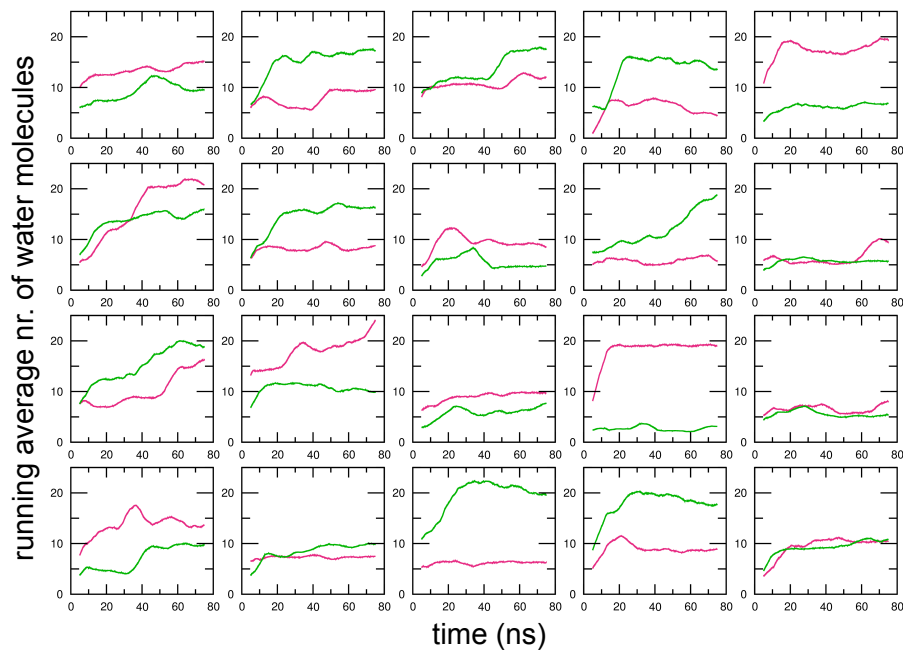


Figure C.1: **Average number of water molecules inside Fluc-Bpe when  $F^-$  is restrained at F82 and F85 (Amber99SB).** Green and magenta curves represent running average number of water molecules inside Fluc-Bpe when  $F^-$  ions are restrained at the F85 and F82 binding sites. Each color corresponds to one pore.

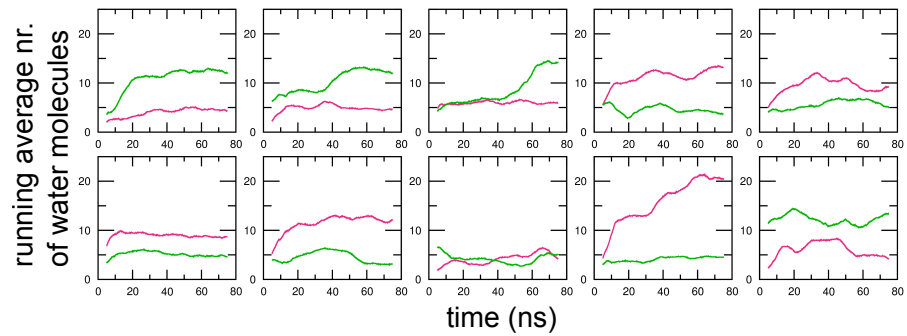


Figure C.2: **Average number of water molecules inside Fluc-Bpe when  $F^-$  is restrained at F82 and F85 (CHARMM36).** Green and magenta curves represent running average number of water molecules inside Fluc-Bpe when  $F^-$  ions are restrained at the F85 and F82 binding sites. Each color corresponds to one pore.

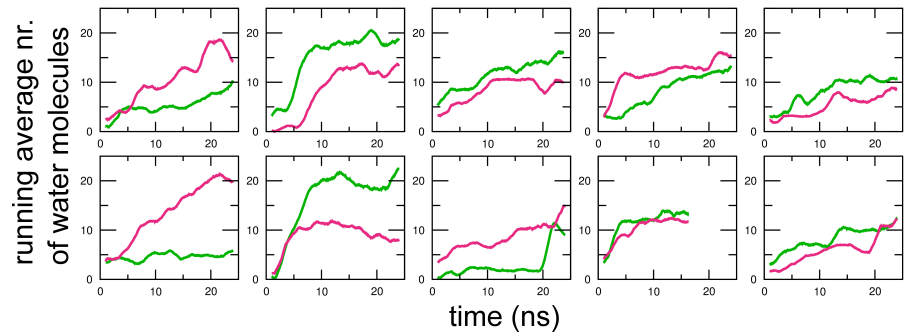


Figure C.3: **Average number of water molecules inside Fluc-Bpe when  $F^-$  is restrained at F82 and F85 (CHARMM-Drude).** Green and magenta curves represent running average number of water molecules inside Fluc-Bpe when  $F^-$  ions are restrained at the F85 and F82 binding sites. Each color corresponds to one pore.

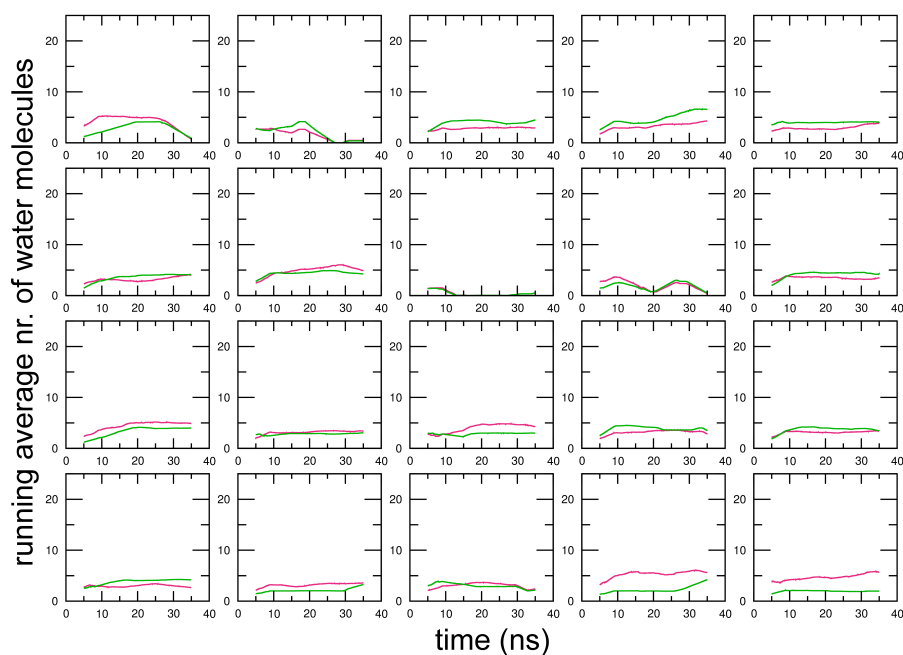


Figure C.4: Average number of water molecules inside Fluc-Bpe when water is bound to F82 and F85. Green and magenta curves represent running average number of water molecules inside Fluc-Bpe when water molecules are bound to the F85 and F82 binding sites. Each color corresponds to one pore.



# D

---

## ABOUT THE LATERAL DISTRIBUTION OF CHOLESTEROL IN BINARY LIPID MIXTURES

---

### D.1 INTRODUCTION

Cholesterol can be found in the vast majority of animal cell membranes. With concentrations up to 50 mol%, the presence of cholesterol in lipidic bilayers increases their order and packing density, which has deep implications in several physicochemical properties, such as the bending moduli, membrane curvature, line tension, lateral diffusion, or permeability [302–306]. Analyses on the lateral distribution of cholesterol in ternary lipid mixtures with saturated and unsaturated lipids have concluded that the domain formation in such membranes is favored by enthalpic contributions of cholesterol in saturated bilayers as compared with unsaturated ones [307–311]. However, in cholesterol-containing binary mixtures, despite the formation of nanoscopic domains in saturated membranes, or the absence of well-defined domains in unsaturated membranes, the lateral organization of cholesterol remains unclear [312–315]. As a consequence, the local cholesterol density fluctuations produced by lateral diffusion are not quantitatively understood. For instance, it is not known whether the enthalpy-favored packing of cholesterol with saturated lipids lead to different degrees of lateral domains. Hence, based on the statistical theory of solutions, we studied the relation between the fluctuations in cholesterol concentration and domain formation with the chemical potential of cholesterol in several binary lipid mixtures. In this work, I analyzed the influence of the chemical potential differences on the formation of cholesterol-depleted domains by means of Coarse-Grained (CG) Molecular Dynamics (MD) simulations of cholesterol in increasingly unsaturated membranes, namely, in membranes of 1,2-dipalmitoyl-*sn*-phosphatidylcholine (DPPC) and 1,2-dioleoyl-*sn*-glycero-3-phosphocholine (DOPC) membranes [316].

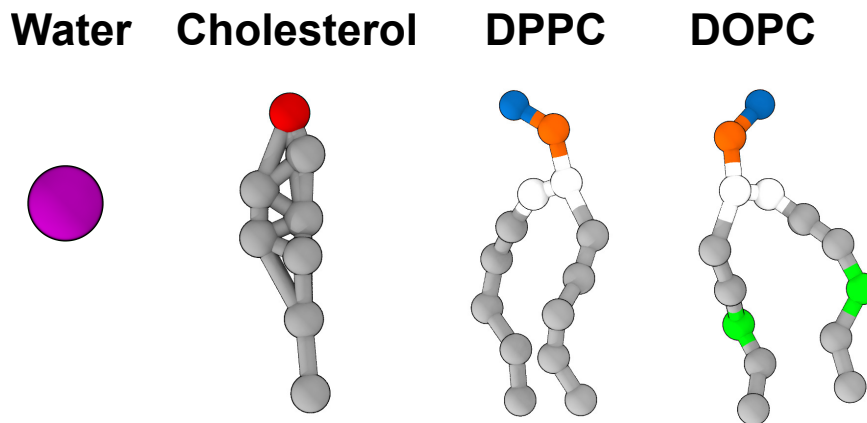


Figure D.1: CG **representation of water and lipids in Martini FF**. Purple: water bead; blue/orange: phospholipid head groups; white: glycerol; silver: hydrocarbon; green: unsaturated groups; red: hydroxyl.

## D.2 MATERIALS AND METHODS

### D.2.1 System Set-up for MD simulations

The simulation systems contained  $\sim 800$  phospholipid-cholesterol molecules and  $\sim 13000$  water beads. The total area of the equilibrated square membranes was between  $198$  and  $263 \text{ nm}^2$ . Molecular interactions were described by the CG Martini force-field [36], which models three to four heavy atoms as one CG bead. The CG representation of the molecules used in this study is showed in Figure D.1. The suitability of this particular force-field for this kind of study is confirmed by the successful simulations of phase separation in ternary lipid mixtures, which reproduced the favoured cholesterol packing with saturated lipids [317, 318].

### D.2.2 Pulling simulation details before US

Once the membrane systems were equilibrated, we carried out pulling simulations along the reaction coordinate  $\xi$  to create a circular cholesterol-depleted domain. The reaction coordinate was taken from Tolpekina *et al.* and is defined as [319]:

$$\xi = \frac{(\Sigma - \Sigma_0)}{(N - \Sigma_0)}, \quad (\text{D.1})$$

where  $\Sigma = \sum_{i=1}^N \tanh(r_i/\zeta)$ , and  $\Sigma_0$  is the equilibrium value of  $\Sigma$ , computed analytically assuming a random distribution of cholesterol in the membrane.  $\zeta = 1.5 \text{ nm}$  specifies the approximate radius of the cholesterol-depleted domain when  $\xi$  is close to one.  $N$  is the number of cholesterol molecules,  $r_i$  is the COM distance in the membrane

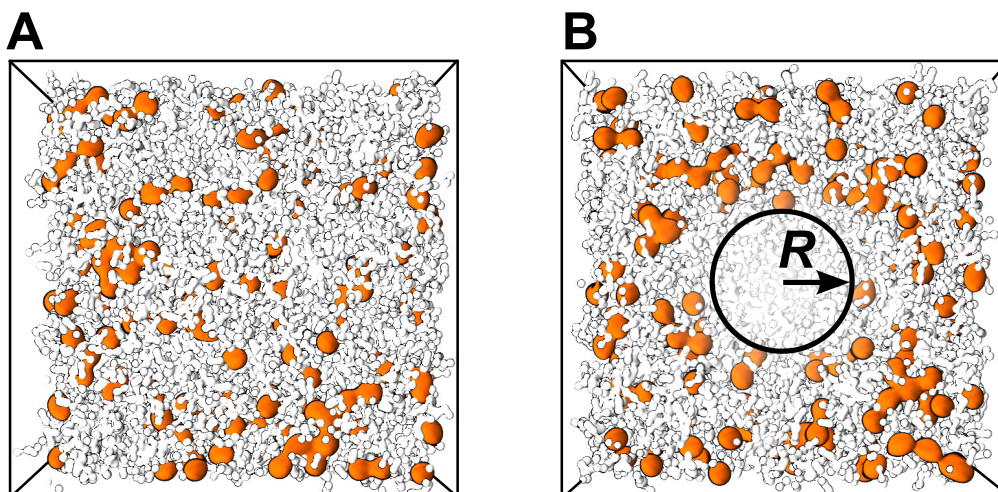


Figure D.2: CG system in pulling simulations. (A) Simulation system of DPPC (white) + 30 mol % cholesterol (orange) in which  $\xi = 0$ . (B) Same simulation system after forming a cholesterol-depleted domain of radius  $R$  ( $\xi = 0.9$ ).

plane of cholesterol molecule  $i$  from the center of the cholesterol-depleted domain. The reaction coordinate  $\xi$  is normalized such that  $\xi = 0$  corresponds to a homogeneous cholesterol distribution, whilst  $\xi \sim 1$  corresponds to the fully formed domain. Figure D.2 shows a typical pulling simulation system before and after pulling along the reaction coordinate  $\xi$ .

### D.2.3 US simulation and PMF calculation details

We computed the PMFs  $W_D(\xi)$  for the formation of a circular cholesterol-depleted domain with US simulations. Accordingly, we generated 21 equally spaced umbrella windows from each pulling simulation. The sampled region of the reaction coordinate was chosen to be between  $\xi = 0$  and  $\xi = 0.9$ , and the force constant of the umbrella potential was  $k = 12500 \text{ kJ mol}^{-1}$ . We simulated each window for  $1 \mu\text{s}$ . The first 40 ns of each trajectory were removed for equilibration. The PMFs were computed using WHAM, and the histograms were bootstrapped for computing statistical errors of  $\leq 1 \text{ kJ mol}^{-1}$  [179, 181]. Since  $\xi$  is little intuitive, we translated the PMFs  $W_D(\xi)$  into functions of the approximate radius  $R$  of the cholesterol-depleted domain, where  $R$  was defined as the radius where cholesterol density reached 50% of the value far away from the domain center, thereby obtaining the PMFs  $W_D(R)$  for domain formation.

#### D.2.4 Loss of lateral entropy calculations

The loss of lateral entropy  $\Delta S_{\text{lat}}(R) = S_{\text{lat}}(R) - S_{\text{lat}}(R = 0)$  upon cholesterol-depleted domain formation was calculated via the Shannon entropy:

$$S_{\text{lat}}(R) = -k_B \int_{\text{box}} dx dy p_R(x, y) \ln p_R(x, y), \quad (\text{D.2})$$

where  $p_R(x, y)$  denotes the lateral density of cholesterol at domain radius  $R$ , taken from the respective US simulations.  $S_{\text{lat}}(R = 0)$  is the lateral entropy in the absence of a domain, and was analytically calculated by assuming a homogeneous cholesterol density  $p_{R=0}(x, y) = N/A_b$ , where  $A_b$  is the area of the membrane. The loss of lateral entropy depends on  $A_b$ , approximately via

$$\Delta S_{\text{lat}}(R) \approx k_B \ln \left( 1 - \frac{\pi R^2}{A_b} \right). \quad (\text{D.3})$$

#### D.2.5 Calculation of $\Delta W_\mu(R)$ contributions

The contribution of the chemical potential to the PMFs for domain formation,  $\Delta W_\mu(R)$ , was calculated in the following way. The number of cholesterol molecules that is transferred from the region of reduced cholesterol content into the region of higher cholesterol content, as a consequence of forming the cholesterol-depleted domain, can be calculated as

$$N_R = \rho \pi R^2 \quad (\text{D.4})$$

where  $\rho = N/A_b$  is the average cholesterol 2D density. Since the cholesterol-depleted domain area is much smaller than the total membrane area, the cholesterol concentration  $x_c$  outside the domain is considered constant. Then, we have

$$\Delta W_\mu(R) = \int_0^{N_R} dN'_R [\Delta \mu(x_c^D(N'_R)) - \Delta \mu(x_c)], \quad (\text{D.5})$$

where  $x_c^D(N'_R) = N'_R x_c / (\rho \pi R^2)$  is the cholesterol concentration inside the domain if  $N'_R$  cholesterol molecules are left. If we approximate the curves in Figure 2D of reference [316] by

$$\Delta \mu(x_c) = m_1 x_c + m_2 x_c^2, \quad (\text{D.6})$$

which yields for the integral

$$\Delta W_\mu(R) = \pi R^2 \rho x_c \left( \frac{1}{2} m_1 + \frac{2}{3} m_2 x_c \right). \quad (\text{D.7})$$



## D.3 RESULTS AND DISCUSSION

The main findings of the study may be summarized as follows [316]:

1. Cholesterol molecules are repelled from cholesterol-enriched domains, which explains why cholesterol does not form such domains in binary mixtures [314].
2. Despite the fact that cholesterol packs more favorably with saturated phospholipids, the excess chemical potential difference  $\Delta\mu_{\text{ex}}$ , i. e. the change in chemical potential with increasing cholesterol fraction, is not influenced by the saturation degree of phospholipids in cholesterol concentrations  $x_c < 40$  mol %.
3. The excess chemical potential  $\mu_{\text{ex}}$ , computed from  $\Pi$ , increases with the number of double bonds in the lipid tails, which is consistent with experimental evidence [311, 320].
4. The lateral distribution of cholesterol is less compressible, i. e. more homogeneous, than that of a 2D ideal gas, as suggested by the computed chemical potential  $\mu(x_c)$ . In turn,  $\mu(x_c)$  can be related to the lateral 2D compressibility of cholesterol  $\chi_{\text{chol}}$  thanks to the statistical theory of solutions.
5. The lower compressibility of the lateral distribution of cholesterol is independent from the type of phospholipid, as  $\Delta\mu_{\text{ex}}$  hardly changes with different lipids.

To aid to understand the effect of the chemical potential differences on the formation of cholesterol-depleted domains, we calculated the PMFs  $W_D(R)$  for cholesterol-depleted domain formation in cholesterol-DPPC and cholesterol-DOPC binary mixtures (see Section D.2).

The PMFs  $W_D(R)$  for cholesterol-depleted domain formation presented in Figure D.3 (solid lines) exhibit a fast increase with increasing radius  $R$ , which indicates that the spontaneous formation of such domains is unlikely. Furthermore, whereas the steepness of the PMFs and the free energy of domain formation increase with increasing concentrations of cholesterol, the profiles show similar trends irrespective of the different degree of saturation of the tested lipids, DPPC and DOPC, which is consistent with the results of the excess chemical potential difference. To relate the differences in the excess chemical potential  $\Delta\mu_{\text{ex}}$  to the PMFs  $W_D(R)$  of domain formation, we can decompose the latter into two components:

- i The lateral entropy  $S_{\text{lat}}$  of cholesterol decreases upon domain formation, which can be calculated via the Shannon entropy (see Section D.2.4). Accordingly, the entropic contribution  $-T\Delta S_{\text{lat}}(R)$  plotted Figure D.3 (dashed lines) was calculated from simulations.

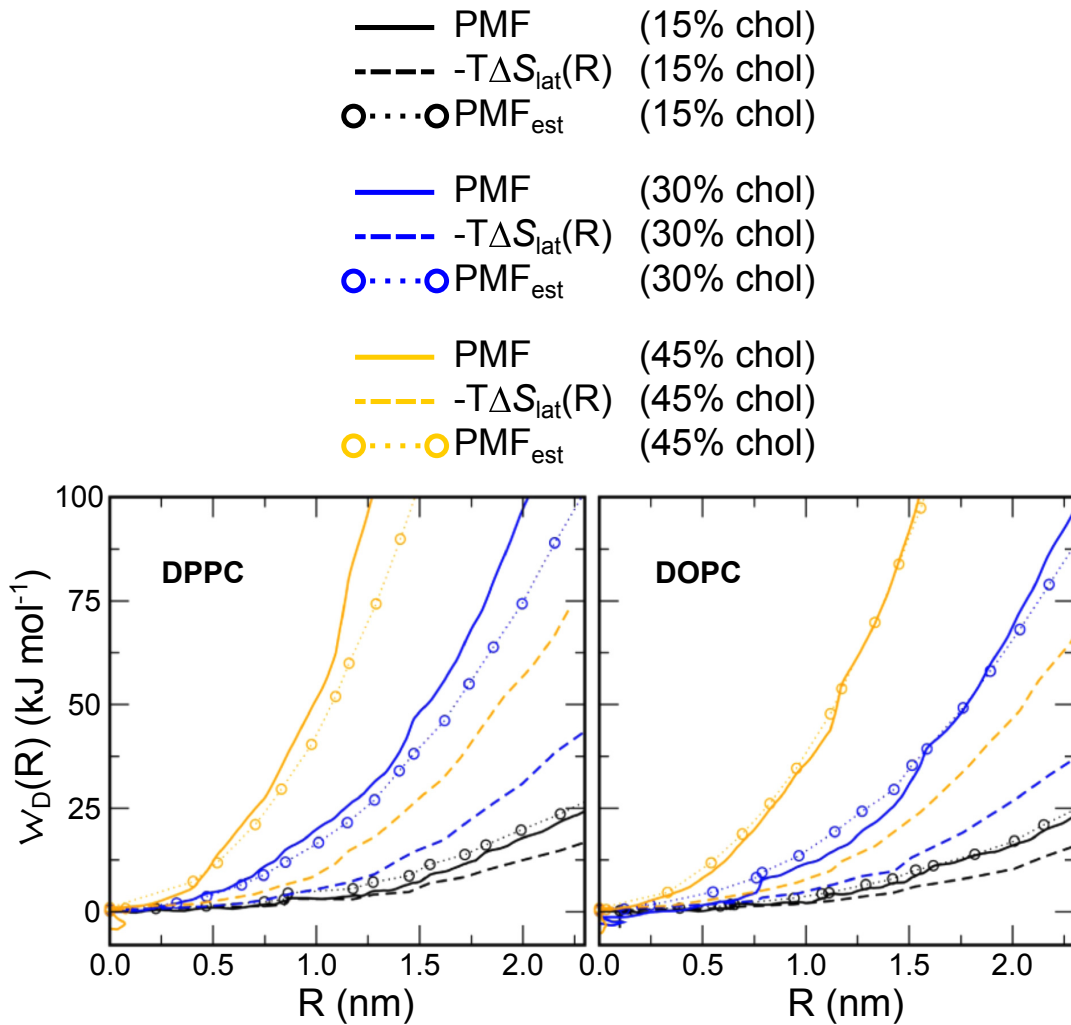


Figure D.3: PMFs  $W_D(R)$  for cholesterol-depleted domain formation. The PMFs for domain formation have been calculated in binary lipid mixtures of cholesterol-DPPC (left) and cholesterol-DOPC (right), with varying concentrations of cholesterol (see legend). Solid lines: PMFs from US simulations; dashed lines: loss of lateral entropy  $-T\Delta S_{\text{lat}}(R)$ ; circles: PMFs estimated via  $\Delta W_{\mu}(R) - T\Delta S_{\text{lat}}(R)$

- ii The contribution of the chemical potential due to the formation of the cholesterol-depleted domain  $\Delta W_{\mu}(R)$ , which can be calculated following the derivation in Section D.2.5. This contribution arises from transferring cholesterol molecules from a cholesterol-depleted domain into a cholesterol-enriched domain.

The sum of the two components,  $\Delta W_{\mu}(R) - T\Delta S_{\text{lat}}(R)$ , is presented in Figure D.3 (circles) labelled as the “estimated”  $\text{PMF}_{\text{est}}$ . It is noteworthy that these  $\text{PMF}_{\text{est}}$  are in excellent agreement with the PMFs  $W_D(R)$  calculated from US simulations (solid lines), which confirms the dependence of the lateral distribution of cholesterol on the combination of i) the lateral entropy  $S_{\text{lat}}$ , and ii) changes in the chemical poten-

tial of cholesterol with the concentration of cholesterol. Furthermore, the balance between these two factors changes with the concentration of cholesterol  $\chi_c$ . Whilst at low  $\chi_c$  the formation of cholesterol-depleted domains is dominated by the lateral entropy (Figure D.3, black lines), similar to a 2D ideal gas, at high  $\chi_c$  the major contribution comes from the repulsive cholesterol-cholesterol interactions (Figure D.3, blue and yellow lines). Overall, these results indicate that the free energy cost for adding a cholesterol molecule increases with packing, and given that tighter packing is associated with lower enthalpy, we suggest that the contribution to that the increasing cost in free energy is mainly entropic. We, therefore, propose that the cholesterol molecules, which are like solvent molecules solvated in a solvent of phospholipids, modify the structure of the solvent, which indirectly repels the incoming cholesterol molecules.

#### D.4 CONCLUSION

In this study we quantified the lateral organization of cholesterol via chemical potentials, which were translated into density fluctuations that are far lower than those of the 2D ideal gas, and into free energies of cholesterol-depleted domain formation. In particular, the results presented in this appendix, i. e. the free energies of domain formation, are in agreement with the chemical potentials and the density fluctuations computed by my colleagues. All together, we found that the lateral distribution of cholesterol in binary lipid mixtures is dominated by the lateral entropy  $S_{\text{lat}}$  only at very low concentrations of cholesterol. In contrast, the cholesterol-cholesterol repulsive interactions dominate at higher concentrations of cholesterol. Besides, we analyzed the influence of unsaturations in the lateral distribution of cholesterol, which were found to play a minor role. This might be non-intuitive, given the preferred packing of cholesterol with saturated lipids [310, 321, 322]. It is important to stress, however, that the packing of cholesterol is related to the magnitude of  $\mu_{\text{ex}}$ , which is strongly influenced by the lipid unsaturations [320, 323, 324], but that the lateral distribution of cholesterol is related to  $\partial\mu_{\text{ex}}/\partial\chi_c$ , which we found to be similar regardless of the lipid unsaturations. We hope that the quantitative analysis carried out in this work helps future researchers to understand better the functional implications of lateral distribution in membranes.



---

## BIBLIOGRAPHY

---

1. Phillips, R., Kondev, J., Theriot, J. & Garcia, H. *Physical biology of the cell* (Garland Science, 2012).
2. Mitra, K., Ubarretxena-Belandia, I., Taguchi, T., Warren, G. & Engelman, D. M. Modulation of the bilayer thickness of exocytic pathway membranes by membrane proteins rather than cholesterol. *Proc. Nat. Acad. Sci.* **101**, 4083–4088 (2004).
3. Berg, J., Stryer, L., Tymoczko, J. & Gatto, G. *Biochemistry* (Macmillan Learning, 2015).
4. Alberts, B, Johnson, A, Lewis, J, Raff, M, Roberts, K & Walter, P. *Molecular Biology of the Cell* (Garland Science, 2008).
5. Ashcroft, F., Gadsby, D. & Miller, C. Introduction. The blurred boundary between channels and transporters. *Philos. Trans. R. Soc. B Biol. Sci.* **364**, 145–147 (2009).
6. Gadsby, D. C. Ion channels versus ion pumps: The principal difference, in principle. *Nat. Rev. Mol. Cell Biol.* **10**, 344–352 (2009).
7. *The Transporter Classification Database* <http://www.tcdb.org/>. Accessed: 2018-02-28.
8. Austin, R. H., Beeson, K.-W., Eisenstein, L, Frauenfelder, H & Gunsalus, I. Dynamics of ligand binding to myoglobin. *Biochemistry* **14**, 5355–5373 (1975).
9. Frauenfelder, H., Sligar, S. G. & Wolynes, P. G. The energy landscapes and motions of proteins. *Science* **254**, 1598–1603 (1991).
10. Jensen, M. R., Zweckstetter, M., Huang, J.-r. & Blackledge, M. Exploring free-energy landscapes of intrinsically disordered proteins at atomic resolution using NMR spectroscopy. *Chem. Rev.* **114**, 6632–6660 (2014).
11. Uversky, V. N. & Dunker, A. K. Understanding protein non-folding. *Biochim. Biophys. Acta* **1804**, 1231–1264 (2010).
12. Boehr, D. D., Nussinov, R. & Wright, P. E. The role of dynamic conformational ensembles in biomolecular recognition. *Nat. Chem. Biol.* **5**, 789 (2009).
13. Tsai, C.-J., Kumar, S., Ma, B. & Nussinov, R. Folding funnels, binding funnels, and protein function. *Prot. Sci.* **8**, 1181–1190 (1999).
14. Ma, B., Shatsky, M., Wolfson, H. J. & Nussinov, R. Multiple diverse ligands binding at a single protein site: A matter of pre-existing populations. *Prot. Sci.* **11**, 184–197 (2002).

15. Tobi, D. & Bahar, I. Structural changes involved in protein binding correlate with intrinsic motions of proteins in the unbound state. *Proc. Natl. Acad. Sci.* **102**, 18908–18913 (2005).
16. Tokuriki, N. & Tawfik, D. S. Protein dynamism and evolvability. *Science* **324**, 203–207 (2009).
17. Henzler-Wildman, K. & Kern, D. Dynamic personalities of proteins. *Nature* **450**, 964–972 (2007).
18. Kai, L., Torchia, D. & Bax, A. Backbone dynamics of proteins as studied by <sup>15</sup>N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry* **28**, 8972 (1989).
19. Baber, J. L., Szabo, A. & Tjandra, N. Analysis of slow interdomain motion of macromolecules using NMR relaxation data. *J. Am. Chem. Soc.* **123**, 3953–3959 (2001).
20. Palmer III, A. G. NMR probes of molecular dynamics: overview and comparison with other techniques. *Annu. Rev. Biophys. Biomol. Struct.* **30**, 129–155 (2001).
21. Wales, T. E. & Engen, J. R. Hydrogen exchange mass spectrometry for the analysis of protein dynamics. *Mass Spect. Rev.* **25**, 158–170 (2006).
22. Parak, F. G. Proteins in action: the physics of structural fluctuations and conformational changes. *Curr. Op. Struct. Biol.* **13**, 552–557 (2003).
23. Baldwin, A. J. & Kay, L. E. NMR spectroscopy brings invisible protein states into focus. *Nature Chem. Biol.* **5**, 808 (2009).
24. Orozco, M. A theoretical view of protein dynamics. *Chem. Soc. Rev.* **43**, 5051–5066 (2014).
25. Karplus, M. & McCammon, J. Dynamics of proteins: elements and function. *Annu. Rev. Biochem.* **52**, 263–300 (1983).
26. McCammon, J. A., Gelin, B. R. & Karplus, M. Dynamics of folded proteins. *Nature* **267**, 585 (1977).
27. Ma, J., Sigler, P. B., Xu, Z. & Karplus, M. A dynamic model for the allosteric mechanism of GroEL<sub>1</sub>. *J. Mol. Biol.* **302**, 303–313 (2000).
28. Tai, K., Shen, T., Börjesson, U., Philippopoulos, M. & McCammon, J. A. Analysis of a 10-ns molecular dynamics simulation of mouse acetylcholinesterase. *Biophys. J.* **81**, 715–724 (2001).
29. Piana, S., Klepeis, J. L. & Shaw, D. E. Assessing the accuracy of physical models used in protein-folding simulations: quantitative evidence from long molecular dynamics simulations. *Curr. Opin. Struct. Biol.* **24**, 98–105 (2014).

30. Huang, J. & MacKerell, A. D. Force field development and simulations of intrinsically disordered proteins. *Curr. Opin. Struct. Biol.* **48**, 40–48 (2018).
31. Lemkul, J. A., Huang, J., Roux, B. & MacKerell Jr, A. D. An empirical polarizable force field based on the classical drude oscillator model: development history and recent applications. *Chem. Rev.* **116**, 4983–5013 (2016).
32. Yu, I., Mori, T., Ando, T., Harada, R., Jung, J., Sugita, Y. & Feig, M. Biomolecular interactions modulate macromolecular structure and dynamics in atomistic model of a bacterial cytoplasm. *eLife* **5**, 1–22 (2016).
33. Shaw, D. E., Dror, R. O., Salmon, J. K., Grossman, J., Mackenzie, K. M., Bank, J. A., Young, C., Deneroff, M. M., Batson, B., Bowers, K. J., *et al.* Millisecond-scale molecular dynamics simulations on Anton in *Proceedings of the conference on high performance computing networking, storage and analysis* (2009), 39.
34. Lane, T. J., Shukla, D., Beauchamp, K. A. & Pande, V. S. To milliseconds and beyond: challenges in the simulation of protein folding. *Curr. Opin. Struct. Biol.* **23**, 58–65 (2013).
35. Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B. & Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1**, 19–25 (2015).
36. Marrink, S. J., Risselada, H. J., Yefimov, S., Tieleman, D. P. & De Vries, A. H. The MARTINI force field: coarse grained model for biomolecular simulations. *J. Phys. Chem. B* **111**, 7812–7824 (2007).
37. Monticelli, L., Kandasamy, S. K., Periole, X., Larson, R. G., Tieleman, D. P. & Marrink, S.-J. The MARTINI coarse-grained force field: extension to proteins. *J. Chem. Theory Comput.* **4**, 819–834 (2008).
38. López, C. A., Rzepiela, A. J., De Vries, A. H., Dijkhuizen, L., Hünenberger, P. H. & Marrink, S. J. Martini coarse-grained force field: extension to carbohydrates. *J. Chem. Theory Comput.* **5**, 3195–3210 (2009).
39. Uusitalo, J. J., Ingólfsson, H. I., Marrink, S. J. & Faustino, I. Martini Coarse-Grained Force Field for RNA. *Biophys. J.* **114**, 437a (2018).
40. *Membrane Proteins of Known 3D Structure* <http://blanco.biomol.uci.edu/mpstruc/>. Accessed: 2018-02-28.
41. Bill, R. M., Henderson, P. J., Iwata, S., Kunji, E. R., Michel, H., Neutze, R., Newstead, S., Poolman, B., Tate, C. G. & Vogel, H. Overcoming barriers to membrane protein structure determination. *Nat. Biotechnol.* **29**, 335 (2011).

42. Mackerell, A. D. Empirical force fields for biological macromolecules: overview and issues. *J. Comput. Chem.* **25**, 1584–1604 (2004).
43. Piggot, T. J., Pineiro, Á. & Khalid, S. Molecular dynamics simulations of phosphatidylcholine membranes: a comparative force field study. *J. Chem. Theory Comput.* **8**, 4593–4609 (2012).
44. Knight, C. J. & Hub, J. S. MemGen: A general web server for the setup of lipid membrane simulation systems. *Bioinformatics* **31**, 2897–2899 (2015).
45. Wolf, M. G., Hoefling, M., Aponte-Santamaría, C., Grubmüller, H. & Groenhof, G. g\_membed: Efficient insertion of a membrane protein into an equilibrated lipid bilayer with minimal perturbation. *J. Comput. Chem.* **31**, 2169–2174 (2010).
46. Kutzner, C., Grubmüller, H., De Groot, B. L. & Zachariae, U. Computational electrophysiology: the molecular dynamics of ion channel permeation and selectivity in atomistic detail. *Biophys. J.* **101**, 809–817 (2011).
47. Dhakshnamoorthy, B., Rohaim, A., Rui, H., Blachowicz, L. & Roux, B. Structural and functional characterization of a calcium-activated cation channel from *Tsukamurella paurometabola*. *Nat. Commun.* **7**, 12753 (2016).
48. Köpfer, D. A., Song, C., Gruene, T., Sheldrick, G. M., Zachariae, U. & de Groot, B. L. Ion permeation in K<sup>+</sup> channels occurs by direct Coulomb knock-on. *Science* **346**, 352–355 (2014).
49. Latorraca, N. R., Fastman, N. M., Venkatakrisnan, A., Frommer, W. B., Dror, R. O. & Feng, L. Mechanism of Substrate Translocation in an Alternating Access Transporter. *Cell* **169**, 96–107 (2017).
50. Hedger, G. & Sansom, M. S. P. Lipid interaction sites on channels, transporters and receptors: Recent insights from molecular dynamics simulations. *Biochim. Biophys. Acta - Biomembr.* 2390–2400 (2016).
51. Kutzner, C., Köpfer, D. A., Machtens, J. P., De Groot, B. L., Song, C. & Zachariae, U. Insights into the function of ion channels by computational electrophysiology simulations. *Biochim. Biophys. Acta - Biomembr.* **1858**, 1741–1752 (2016).
52. Joh, N. H., Wang, T., Bhate, M. P., Acharya, R., Wu, Y., Grabe, M., Hong, M., Grigoryan, G. & DeGrado, W. F. De novo design of a transmembrane Zn<sup>2+</sup>-transporting four-helix bundle. *Science* **346**, 1520–1524 (2014).
53. Mannuzzu, L. M., Moronne, M. M. & Macey, R. I. Estimate of the number of urea transport sites in erythrocyte ghosts using a hydrophobic mercurial. *J. Mem. Biol.* **133**, 85–97 (1993).



54. MacIver, B., Smith, C. P., Hill, W. G. & Zeidel, M. L. Functional characterization of mouse urea transporters UT-A2 and UT-A3 expressed in purified *Xenopus laevis* oocyte plasma membranes. *Am. J. Physiol. Renal Physiol.* **294**, F956–F964 (2008).
55. Mistry, A. C., Chen, G., Kato, A., Nag, K., Sands, J. M. & Hirose, S. A novel type of urea transporter, UT-C, is highly expressed in proximal tubule of seawater eel kidney. *Am. J. Physiol. Renal Physiol.* **288**, F455–F465 (2005).
56. Fenton, R. A., Cooper, G. J., Morris, I. D. & Smith, C. P. Coordinated expression of UT-A and UT-B urea transporters in rat testis. *Am. J. Physiol. Cell Physiol.* **282**, C1492–C1501 (2002).
57. Trinh-Trang-Tan, M.-M., Lasbennes, F., Gane, P., Roudier, N., Ripoche, P., Cartron, J.-P. & Bailly, P. UT-B1 proteins in rat: tissue distribution and regulation by antidiuretic hormone in kidney. *Am. J. Physiol. Renal Physiol.* **283**, F912–F922 (2002).
58. Kwun, Y.-S., Yeo, S. W., Ahn, Y.-H., Lim, S.-W., Jung, J.-Y., Kim, W.-Y., Sands, J. M. & Kim, J. Immunohistochemical localization of urea transporters A and B in the rat cochlea. *Hear. Res.* **183**, 84–96 (2003).
59. Coyle, J., McDaid, S., Walpole, C & Stewart, G. S. UT-B urea transporter localization in the bovine gastrointestinal tract. *J. Mem. Biol.* **249**, 77–85 (2016).
60. Hou, R., Alemozaffar, M., Yang, B., Sands, J. M., Kong, X. & Chen, G. Identification of a Novel UT-B Urea Transporter in Human Urothelial Cancer. *Frontiers in physiology* **8**, 245 (2017).
61. Yang, B., Bankir, L., Gillespie, A., Epstein, C. J. & Verkman, A. Urea-selective concentrating defect in transgenic mice lacking urea transporter UT-B. *J. Biol. Chem.* **277**, 10633–10637 (2002).
62. Yang, B. & Verkman, A. Analysis of Double Knockout Mice Lacking Aquaporin-1 and Urea Transporter UT-B EVIDENCE FOR UT-B-FACILITATED WATER TRANSPORT IN ERYTHROCYTES. *J. Biol. Chem.* **277**, 36782–36786 (2002).
63. Geyer, R. R., Musa-Aziz, R., Enkavi, G., Mahinthichaichan, P., Tajkhorshid, E. & Boron, W. F. Movement of NH<sub>3</sub> through the human urea transporter B: a new gas channel. *Am. J. Physiol. Renal Physiol.* **304**, F1447–F1457 (2013).
64. Levin, E. J., Cao, Y., Enkavi, G., Quick, M., Pan, Y., Tajkhorshid, E. & Zhou, M. Structure and permeation mechanism of a mammalian urea transporter. *Proc. Natl. Acad. Sci.* **109**, 11194–11199 (2012).
65. Ariz-Extreme, I. & Hub, J. S. Potential of mean force calculations of solute permeation across UT-B and AQP1: A comparison between molecular dynamics and 3D-RISM. *J. Phys. Chem. B* **121**, 1506–1519 (2017).

66. Weinstein, L. H. & Davison, A. *Fluorides in the environment: effects on plants and animals* (CABI, 2004).
67. Breaker, R. New insight on the response of bacteria to fluoride. *Caries Res.* **46**, 78–81 (2012).
68. Marquis, R. E., Clock, S. A. & Mota-Meira, M. Fluoride and organic weak acids as modulators of microbial physiology. *FEMS Microbiol. Rev.* **26**, 493–510 (2003).
69. Barbier, O., Arreola-Mendoza, L. & Del Razo, L. M. Molecular mechanisms of fluoride toxicity. *Chem. Biol. Inter.* **188**, 319–333 (2010).
70. Ji, C., Stockbridge, R. B. & Miller, C. Bacterial fluoride resistance, Fluc channels, and the weak acid accumulation effect. *J. Gen. Physiol.* **144**, 257–261 (2014).
71. Accardi, A. & Picollo, A. CLC channels and transporters: proteins with borderline personalities. *Biochim. Biophys. Acta* **1798**, 1457–1464 (2010).
72. Stockbridge, R. B., Lim, H.-H., Otten, R., Williams, C., Shane, T., Weinberg, Z. & Miller, C. Fluoride resistance and transport by riboswitch-controlled CLC antiporters. *Proc. Natl. Acad. Sci.* **109**, 15289–15294 (2012).
73. Brammer, A. E., Stockbridge, R. B. & Miller, C. F<sup>-</sup>/Cl<sup>-</sup> selectivity in CLCF-type F<sup>-</sup>/H<sup>+</sup> antiporters. *J. Gen. Physiol.* **144**, 129–136 (2014).
74. Stockbridge, R. B., Robertson, J. L., Kolmakova-Partensky, L. & Miller, C. A family of fluoride-specific ion channels with dual-topology architecture. *eLife* **2** (2013).
75. Li, S., Smith, K. D., Davis, J. H., Gordon, P. B., Breaker, R. R. & Strobel, S. A. Eukaryotic resistance to fluoride toxicity mediated by a widespread family of fluoride export proteins. *Proc. Natl. Acad. Sci.* **110**, 19018–19023 (2013).
76. Berbasova, T., Nallur, S., Sells, T., Smith, K. D., Gordon, P. B., Tausta, S. L. & Strobel, S. A. Fluoride export (FEX) proteins from fungi, plants and animals are 'single barreled' channels containing one functional and one vestigial ion pore. *PloS One* **12**, e0177096 (2017).
77. Baker, J. L., Sudarsan, N., Weinberg, Z., Roth, A., Stockbridge, R. B. & Breaker, R. R. Widespread genetic switches and toxicity resistance proteins for fluoride. *Science* **335**, 233–235 (2012).
78. Speed, M. C., Burkhart, B. W., Picking, J. W. & Santangelo, T. J. An archaeal, fluoride-responsive riboswitch provides an inducible expression system for hyperthermophiles. *Appl. Environ. Microbiol.* AEM-02306 (2018).

79. Ren, A., Rajashankar, K. R. & Patel, D. J. Fluoride ion encapsulation by Mg<sup>2+</sup> ions and phosphates in a fluoride riboswitch. *Nature* **486**, 85 (2012).
80. Gälli, R. & Leisinger, T. Specialized bacterial strains for the removal of dichloromethane from industrial waste. *Conserv. Recy.* **8**, 91–100 (1985).
81. Vuilleumier, S., Sorribas, H. & Leisinger, T. Identification of a novel determinant of glutathione affinity in dichloromethane dehalogenases/glutathioneS-transferases. *Biochem. Biophys. Res. Commun.* **238**, 452–456 (1997).
82. Ochsner, A. M., Sonntag, F., Buchhaupt, M., Schrader, J. & Vorholt, J. A. *Methylobacterium extorquens*: methylotrophy and biotechnological applications. *Appl. Microbiol. Biotechnol.* **99**, 517–534 (2015).
83. Macdonald, C. B. & Stockbridge, R. B. A topologically diverse family of fluoride channels. *Curr. Op. Struct. Biol.* **45**, 142–149 (2017).
84. Smith, K. D., Gordon, P. B., Rivetta, A., Allen, K. E., Berbasova, T., Slayman, C. & Strobel, S. A. Yeast Fex1p is a constitutively expressed fluoride channel with functional asymmetry of its two homologous domains. *J. Biol. Chem.* **290**, 19874–19887 (2015).
85. Stockbridge, R. B., Kolmakova-Partensky, L., Shane, T., Koide, A., Koide, S., Miller, C. & Newstead, S. Crystal structures of a double-barrelled fluoride ion channel. *Nature* **525**, 548 (2015).
86. Stockbridge, R. B., Koide, A., Miller, C. & Koide, S. Proof of dual-topology architecture of Fluc F- channels with monobody blockers. *Nature Comm.* **5**, 5120 (2014).
87. Turman, D. L., Nathanson, J. T., Stockbridge, R. B., Street, T. O. & Miller, C. Two-sided block of a dual-topology F- channel. *Proc. Natl. Acad. Sci.* **112**, 5697–5701 (2015).
88. Turman, D. L. & Stockbridge, R. B. Mechanism of single- and double-sided inhibition of dual topology fluoride channels by synthetic monobodies. *J. Gen. Physiol.* jgp–201611747 (2017).
89. Ubarretxena-Belandia, I., Baldwin, J. M., Schuldiner, S. & Tate, C. G. Three-dimensional structure of the bacterial multidrug transporter EmrE shows it is an asymmetric homodimer. *EMBO J.* **22**, 6175–6181 (2003).
90. Morrison, E. A., DeKoster, G. T., Dutta, S., Vafabakhsh, R., Clarkson, M. W., Bahl, A., Kern, D., Ha, T. & Henzler-Wildman, K. A. Antiparallel EmrE exports drugs by exchanging between asymmetric structures. *Nature* **481**, 45 (2012).

91. Ariz-Extreme, I. & Hub, J. S. Assigning crystallographic electron densities with free energy calculations—The case of the fluoride channel Fluc. *PLoS one* **13**, e0196751 (2018).
92. Meyer, E. A., Castellano, R. K. & Diederich, F. Interactions with aromatic rings in chemical and biological recognition. *Angew. Chem. Int. Ed.* **42**, 1210–1250 (2003).
93. Jackson, M. R., Beahm, R., Duvvuru, S., Narasimhan, C., Wu, J., Wang, H.-N., Philip, V. M., Hinde, R. J. & Howell, E. E. A preference for edgewise interactions between aromatic rings and carboxylate anions: The biological relevance of anion- quadrupole interactions. *J. Phys. Chem. B* **111**, 8242–8249 (2007).
94. Philip, V., Harris, J., Adams, R., Nguyen, D., Spiers, J., Baudry, J., Howell, E. E. & Hinde, R. J. A survey of aspartate- phenylalanine and glutamate- phenylalanine interactions in the Protein Data Bank: searching for anion-  $\pi$  pairs. *Biochemistry* **50**, 2939–2950 (2011).
95. Hille, B. & Schwarz, W. Potassium channels as multi-ion single-file pores. *J. Gen. Physiol.* **72**, 409–442 (1978).
96. Hummer, G., Pratt, L. R. & Garcia, A. E. Free energy of ionic hydration. *J. Phys. Chem.* **100**, 1206–1215 (1996).
97. Last, N. B., Kolmakova-Partensky, L., Shane, T. & Miller, C. Mechanistic signs of double-barreled structure in a fluoride ion channel. *eLife* **5** (2016).
98. Last, N. B., Sun, S., Pham, M. C. & Miller, C. Molecular determinants of permeation in a fluoride-specific ion channel. *eLife* **6** (2017).
99. Phongphanphanee, S., Yoshida, N. & Hirata, F. Molecular selectivity in aquaporin channels studied by the 3D-RISM theory. *J. Phys. Chem. B* **114**, 7967–73. ISSN: 1520-5207 (2010).
100. Karplus, M. & McCammon, J. A. Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol.* **9**, 646 (2002).
101. Berendsen, H. J. *Simulating the physical world: hierarchical modeling from quantum mechanics to fluid dynamics* (Cambridge University Press, 2007).
102. Tully, J. C. Perspective on “zur quantentheorie der molekeln”. *Theor. Chem. Acc.* **103**, 173–176 (2000).
103. Arndt, M., Juffmann, T. & Vedral, V. Quantum physics meets biology. *HFSP J.* **3**, 386–400 (2009).
104. Kiefer, C. Emergence of a classical Universe from quantum gravity and cosmology. *Phil. Trans. R. Soc. A* **370**, 4566–4575 (2012).
105. Karplus, M. & Petsko, G. A. Molecular dynamics simulations in biology. *Nature* **347**, 631 (1990).

106. Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W. & Kollman, P. A. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **117**, 5179–5197 (1995).
107. Wang, J., Cieplak, P. & Kollman, P. A. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.* **21**, 1049–1074 (2000).
108. Duan, Y., Wu, C., Chowdhury, S., Lee, M. C., Xiong, G., Zhang, W., Yang, R., Cieplak, P., Luo, R., Lee, T., *et al.* A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* **24**, 1999–2012 (2003).
109. Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J. L., Dror, R. O. & Shaw, D. E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins: Struct., Funct., Bioinform.* **78**, 1950–1958 (2010).
110. Pérez, A., Marchán, I., Svozil, D., Sponer, J., Cheatham III, T. E., Laughton, C. A. & Orozco, M. Refinement of the AMBER force field for nucleic acids: improving the description of  $\alpha/\gamma$  conformers. *Biophys. J.* **92**, 3817–3829 (2007).
111. Banáš, P., Hollas, D., Zgarbová, M., Jurecka, P., Orozco, M., Cheatham III, T. E., Šponer, J. & Otyepka, M. Performance of molecular mechanics force fields for RNA simulations: stability of UUCG and GNRA hairpins. *J. Chem. Theory Comput.* **6**, 3836–3849 (2010).
112. Krepl, M., Zgarbová, M., Stadlbauer, P., Otyepka, M., Banáš, P., Koca, J., Cheatham III, T. E., Jurecka, P. & Šponer, J. Reference simulations of noncanonical nucleic acids with different  $\chi$  variants of the AMBER force field: quadruplex DNA, quadruplex RNA, and Z-DNA. *J. Chem. Theory Comput.* **8**, 2506–2520 (2012).
113. Woods, R. J., Dwek, R. A., Edge, C. J. & Fraser-Reid, B. Molecular mechanical and molecular dynamic simulations of glycoproteins and oligosaccharides. 1. GLYCAM\_93 parameter development. *J. Phys. Chem.* **99**, 3832–3846 (1995).
114. Basma, M., Sundara, S., Çalgan, D., Vernali, T. & Woods, R. J. Solvated ensemble averaging in the calculation of partial atomic charges. *J. Comput. Chem.* **22**, 1125–1137 (2001).
115. Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A. & Simmerling, C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Struct., Funct., Bioinform.* **65**, 712–725 (2006).

116. Dickson, C. J., Rosso, L., Betz, R. M., Walker, R. C. & Gould, I. R. GAFFlipid: a General Amber Force Field for the accurate molecular dynamics simulation of phospholipid. *Soft Matter* **8**, 9617–9627 (2012).
117. MacKerell Jr, A. D., Bashford, D., Bellott, M., Dunbrack Jr, R. L., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H, Ha, S., *et al.* All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **102**, 3586–3616 (1998).
118. Foloppe, N. & MacKerell Jr, A. D. All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data. *J. Comput. Chem.* **21**, 86–104 (2000).
119. Mackerell, A. D. & Banavali, N. K. All-atom empirical force field for nucleic acids: II. Application to molecular dynamics simulations of DNA and RNA in solution. *J. Comput. Chem.* **21**, 105–120 (2000).
120. Hart, K., Foloppe, N., Baker, C. M., Denning, E. J., Nilsson, L. & MacKerell Jr, A. D. Optimization of the CHARMM additive force field for DNA: Improved treatment of the BI/BII conformational equilibrium. *J. Chem. Theory Comput.* **8**, 348–362 (2011).
121. Denning, E. J., Priyakumar, U, Nilsson, L. & Mackerell, A. D. Impact of 2-hydroxyl sampling on the conformational properties of RNA: Update of the CHARMM all-atom additive force field for RNA. *J. Comput. Chem.* **32**, 1929–1943 (2011).
122. Best, R. B., Zhu, X., Shim, J., Lopes, P. E., Mittal, J., Feig, M. & MacKerell Jr, A. D. 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. *J. Chem. Theory Comput.* **8**, 3257–3273 (2012).
123. Pastor, R. & MacKerell Jr, A. Development of the CHARMM force field for lipids. *J. Phys. Chem. Lett.* **2**, 1526–1532 (2011).
124. Guvench, O., Greene, S. N., Kamath, G., Brady, J. W., Venable, R. M., Pastor, R. W. & Mackerell, A. D. Additive empirical force field for hexopyranose monosaccharides. *J. Comput. Chem.* **29**, 2543–2564 (2008).
125. Hatcher, E. R., Guvench, O. & MacKerell Jr, A. D. CHARMM additive all-atom force field for acyclic polyalcohols, acyclic carbohydrates, and inositol. *J. Chem. Theory Comput.* **5**, 1315–1327 (2009).
126. Raman, E. P., Guvench, O. & MacKerell Jr, A. D. CHARMM additive all-atom force field for glycosidic linkages in carbohydrates involving furanoses. *J. Phys. Chem. B* **114**, 12981–12994 (2010).

127. Lee, S., Tran, A., Allsopp, M., Lim, J. B., Héning, J. & Klauda, J. B. CHARMM36 united atom chain model for lipids and surfactants. *J. Phys. Chem. B* **118**, 547–556 (2014).
128. Oostenbrink, C., Villa, A., Mark, A. E. & Van Gunsteren, W. F. A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. *J. Comput. Chem.* **25**, 1656–1676 (2004).
129. Soares, T. A., Hünenberger, P. H., Kastenholtz, M. A., Kräutler, V., Lenz, T., Lins, R. D., Oostenbrink, C. & van Gunsteren, W. F. An improved nucleic acid parameter set for the GROMOS force field. *J. Comput. Chem.* **26**, 725–737 (2005).
130. Schmid, N., Eichenberger, A. P., Choutko, A., Riniker, S., Winger, M., Mark, A. E. & van Gunsteren, W. F. Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *Eur. Biophys. J.* **40**, 843 (2011).
131. Horta, B. A., Fuchs, P. F., van Gunsteren, W. F. & Hünenberger, P. H. New interaction parameters for oxygen compounds in the GROMOS force field: Improved pure-liquid and solvation properties for alcohols, ethers, aldehydes, ketones, carboxylic acids, and esters. *J. Chem. Theory Comput.* **7**, 1016–1031 (2011).
132. Pol-Fachin, L., Rusu, V. H., Verli, H. & Lins, R. D. GROMOS 53A6GLYC, an improved GROMOS force field for hexopyranose-based carbohydrates. *J. Chem. Theory Comput.* **8**, 4681–4690 (2012).
133. Jorgensen, W. L., Maxwell, D. S. & Tirado-Rives, J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* **118**, 11225–11236 (1996).
134. Damm, W., Frontera, A., Tirado-Rives, J. & Jorgensen, W. L. OPLS all-atom force field for carbohydrates. *J. Comput. Chem.* **18**, 1955–1970 (1997).
135. Kaminski, G. A., Friesner, R. A., Tirado-Rives, J. & Jorgensen, W. L. Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J. Phys. Chem. B* **105**, 6474–6487 (2001).
136. Kony, D., Damm, W., Stoll, S. & Van Gunsteren, W. F. An improved OPLS-AA force field for carbohydrates. *J. Comput. Chem.* **23**, 1416–1429 (2002).
137. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **25**, 1157–1174 (2004).

138. Vanommeslaeghe, K., Hatcher, E., Acharya, C., Kundu, S., Zhong, S., Shim, J., Darian, E., Guvench, O., Lopes, P., Vorobyov, I., *et al.* CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* **31**, 671–690 (2010).
139. Lopes, P. E., Roux, B. & MacKerell, A. D. Molecular modeling and dynamics studies with explicit inclusion of electronic polarizability: theory and applications. *Theor. Chem. Acc.* **124**, 11–28 (2009).
140. Cieplak, P., Dupradeau, F.-Y., Duan, Y. & Wang, J. Polarization effects in molecular mechanical force fields. *J. Phys.: Condens. Matter* **21**, 333102 (2009).
141. Kaminski, G. A., Stern, H. A., Berne, B. J., Friesner, R. A., Cao, Y. X., Murphy, R. B., Zhou, R. & Halgren, T. A. Development of a polarizable force field for proteins via ab initio quantum chemistry: first generation model and gas phase tests. *J. Comput. Chem.* **23**, 1515–1531 (2002).
142. Shi, Y., Xia, Z., Zhang, J., Best, R., Wu, C., Ponder, J. W. & Ren, P. Polarizable atomic multipole-based AMOEBA force field for proteins. *J. Chem. Theory Comput.* **9**, 4046–4063 (2013).
143. Ponder, J. W., Wu, C., Ren, P., Pande, V. S., Chodera, J. D., Schnieders, M. J., Haque, I., Mobley, D. L., Lambrecht, D. S., DiStasio Jr, R. A., *et al.* Current status of the AMOEBA polarizable force field. *J. Phys. Chem. B* **114**, 2549–2564 (2010).
144. Rick, S. W., Stuart, S. J. & Berne, B. J. Dynamical fluctuating charge force fields: Application to liquid water. *J. Chem. Phys.* **101**, 6141–6156 (1994).
145. Patel, S. & Brooks, C. L. CHARMM fluctuating charge force field for proteins: I parameterization and application to bulk organic liquid simulations. *J. Comput. Chem.* **25**, 1–16 (2004).
146. Patel, S., Mackerell, A. D. & Brooks, C. L. CHARMM fluctuating charge force field for proteins: II protein/solvent properties from molecular dynamics simulations using a nonadditive electrostatic model. *J. Comput. Chem.* **25**, 1504–1514 (2004).
147. Lamoureux, G., MacKerell Jr, A. D. & Roux, B. A simple polarizable model of water based on classical Drude oscillators. *J. Chem. Phys.* **119**, 5185–5197 (2003).
148. Lamoureux, G. & Roux, B. Modeling induced polarization with classical Drude oscillators: Theory and molecular dynamics simulation algorithm. *J. Chem. Phys.* **119**, 3025–3039 (2003).
149. Thole, B. T. Molecular polarizabilities calculated with a modified dipole interaction. *Chem. Phys.* **59**, 341–350 (1981).



150. Berendsen, H. J., van der Spoel, D. & van Drunen, R. GRO-MACS: a message-passing parallel molecular dynamics implementation. *Comput. Phys. Comm.* **91**, 43–56 (1995).
151. Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E. & Berendsen, H. J. GROMACS: fast, flexible, and free. *J. Comput. Chem.* **26**, 1701–1718 (2005).
152. Pronk, S., Páll, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., Shirts, M. R., Smith, J. C., Kasson, P. M., Van Der Spoel, D., *et al.* GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **29**, 845–854 (2013).
153. Abraham, M., Van Der Spoel, D, Lindahl, E, Hess, B, van Buuren, A., Apol, E, Meulenhoff, P., Tieleman, D., Sijbers, A., Feenstra, K., *et al.* *GROMACS User Manual version 5.1* 2015.
154. Hockney, R., Goel, S. & Eastwood, J. Quiet high-resolution computer models of a plasma. *J. Comput. Phys.* **14**, 148–158 (1974).
155. Miyamoto, S. & Kollman, P. A. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* **13**, 952–962 (1992).
156. Hess, B., Bekker, H., Berendsen, H. J., Fraaije, J. G., *et al.* LINCS: a linear constraint solver for molecular simulations. *J. Comput. Chem.* **18**, 1463–1472 (1997).
157. Hess, B. P-LINCS: A parallel linear constraint solver for molecular simulation. *J. Chem. Theory Comput.* **4**, 116–122 (2008).
158. Feenstra, K. A., Hess, B. & Berendsen, H. J. Improving efficiency of large timescale molecular dynamics simulations of hydrogen-rich systems. *J. Comput. Chem.* **20**, 786–798 (1999).
159. Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S & Karplus, M. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **4**, 187–217 (1983).
160. Saito, M. Molecular dynamics simulations of proteins in solution: artifacts caused by the cutoff approximation. *J. Chem. Phys.* **101**, 4055–4061 (1994).
161. Saito, M. Molecular dynamics/free energy study of a protein in solution with all degrees of freedom and long-range Coulomb interactions. *J. Phys. Chem.* **99**, 17043–17048 (1995).
162. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An N log (N) method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089–10092 (1993).
163. Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H. & Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **103**, 8577–8593 (1995).

164. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126**, 014101 (2007).
165. Berendsen, H. J., Postma, J. v., van Gunsteren, W. F., DiNola, A. & Haak, J. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684–3690 (1984).
166. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. App. Phys.* **52**, 7182–7190 (1981).
167. Nosé, S. & Klein, M. Constant pressure molecular dynamics for molecular systems. *Mol. Phys.* **50**, 1055–1076 (1983).
168. Lipparini, F., Lagardère, L., Stamm, B., Cancès, E., Schnieders, M., Ren, P., Maday, Y. & Piquemal, J.-P. Scalable evaluation of polarization energy and associated forces in polarizable molecular dynamics: I. toward massively parallel direct space computations. *J. Chem. Theory Comput.* **10**, 1638–1651 (2014).
169. Kirkwood, J. G. Statistical mechanics of fluid mixtures. *J. Chem. Phys.* **3**, 300–313 (1935).
170. Roux, B. The calculation of the potential of mean force using computer simulations. *Comput. Phys. Comm.* **91**, 275–282 (1995).
171. Zuckerman, D. M. *Statistical Physics of Biomolecules: An Introduction* (CRC Press, 2010).
172. Hub, J. S. Selectivity, Regulation, and Inhibition of Aquaporin Channels. A Molecular Dynamics Study. *Ph.D. Thesis, Georg-August-Universität Göttingen* (2008).
173. *Statistical Biophysics Blog* <http://statisticalbiophysicsblog.org/>. Accessed: 2018-03-23.
174. Kosztin, I., Barz, B. & Janosi, L. Calculating potentials of mean force and diffusion coefficients from nonequilibrium processes without Jarzynski's equality. *J. Chem. Phys.* **124**, 064106 (2006).
175. Grossfield, A. & Zuckerman, D. M. Quantifying uncertainty and sampling quality in biomolecular simulations. *Ann. Rep. Comput. Chem.* **5**, 23–48 (2009).
176. McGibbon, R. T., Husic, B. E. & Pande, V. S. Identification of simple reaction coordinates from complex dynamics. *J. Chem. Phys.* **146**, 044109 (2017).
177. Torrie, G. M. & Valleau, J. P. Monte Carlo free energy estimates using non-Boltzmann sampling: Application to the sub-critical Lennard-Jones fluid. *Chem. Phys. Lett.* **28**, 578–581 (1974).
178. Kästner, J. Umbrella sampling. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **1**, 932–942 (2011).

179. Kumar, S., Rosenberg, J. M., Bouzida, D., Swendsen, R. H. & Kollman, P. A. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* **13**, 1011–1021 (1992).
180. Kästner, J. & Thiel, W. Bridging the gap between thermodynamic integration and umbrella sampling provides a novel analysis method: “Umbrella integration”. *J. Chem. Phys.* **123**, 144104 (2005).
181. Hub, J. S., De Groot, B. L. & Van Der Spoel, D. g\_wham — A Free Weighted Histogram Analysis Implementation Including Robust Error and Autocorrelation Estimates. *J. Chem. Theory Comput.* **6**, 3713–3720 (2010).
182. Straatsma, T. & Berendsen, H. Free energy of ionic hydration: Analysis of a thermodynamic integration technique to evaluate free energy differences by molecular dynamics simulations. *J. Chem. Phys.* **89**, 5876–5886 (1988).
183. Beveridge, D. L. & DiCapua, F. Free energy via molecular simulation: applications to chemical and biomolecular systems. *Ann. Rev. Biophys. Biophys. Chem.* **18**, 431–492 (1989).
184. Shirts, M. R., Mobley, D. L. & Chodera, J. D. Alchemical free energy calculations: ready for prime time? *Ann. Rep. Comput. Chem.* **3**, 41–59 (2007).
185. Michel, J. & Essex, J. W. Prediction of protein–ligand binding affinity by free energy simulations: assumptions, pitfalls and expectations. *J. Comput. Aided Mol. Des.* **24**, 639–658 (2010).
186. Christ, C. D., Mark, A. E. & Van Gunsteren, W. F. Basic ingredients of free energy calculations: a review. *J. Comput. Chem.* **31**, 1569–1582 (2010).
187. Shirts, M. R. & Mobley, D. L. in *Biomol. Sim.* 271–311 (Springer, 2013).
188. Chodera, J. D., Mobley, D. L., Shirts, M. R., Dixon, R. W., Branson, K. & Pande, V. S. Alchemical free energy methods for drug discovery: progress and challenges. *Curr. Op. Struct. Biol.* **21**, 150–160 (2011).
189. Deng, Y. & Roux, B. Computations of standard binding free energies with molecular dynamics simulations. *J. Phys. Chem. B* **113**, 2234–2246 (2009).
190. Hirata, F. *Molecular Theory of Solvation* (ed Hirata, F.) ISBN: 1-4020-1562-3 (Kluwer Academic Publishers, Dordrecht, 2004).
191. Kovalenko, A. & Hirata, F. Three-dimensional density profiles of water in contact with a solute of arbitrary shape: a RISM approach. *Chem. Phys. Lett.* **290**, 237–244. ISSN: 00092614 (1998).

192. Kovalenko, A. & Hirata, F. Self-consistent description of a metal–water interface by the Kohn–Sham density functional theory and the three-dimensional reference interaction site model. *J. Chem. Phys.* **110**, 10095–10112. ISSN: 00219606 (1999).
193. Hansen, J.-P. & McDonald, I. R. *Theory of Simple Liquids: With Applications to Soft Matter* (Academic Press, 2013).
194. Perkyuns, J. & Pettitt, B. M. A site–site theory for finite concentration saline solutions. *J. Chem. Phys.* **97**, 7656–7666 (1992).
195. Perkyuns, J. & Pettitt, B. M. A dielectrically consistent interaction site theory for solvent–electrolyte mixtures. *Chem. Phys. Lett.* **190**, 626–630 (1992).
196. Baxter, R. Ornstein–Zernike relation and Percus–Yevick approximation for fluid mixtures. *J. Chem. Phys.* **52**, 4559–4562 (1970).
197. Waisman, E. & Lebowitz, J. L. Mean spherical model integral equation for charged hard spheres I. Method of solution. *J. Chem. Phys.* **56**, 3086–3093 (1972).
198. Luchko, T., Gusarov, S., Roe, D. R., Simmerling, C., Case, D. a., Tuszynski, J. & Kovalenko, A. Three-dimensional molecular theory of solvation coupled with molecular dynamics in Amber. *J. Chem. Theory Comput.* **6**, 607–624. ISSN: 1549-9618 (2010).
199. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. The Protein Data Bank, 235–242 (2000).
200. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983).
201. Berger, O., Edholm, O. & Jähnig, F. Molecular dynamics simulations of a fluid bilayer of dipalmitoylphosphatidylcholine at full hydration, constant pressure, and constant temperature. *Biophys. J.* **72**, 2002–2013 (1997).
202. Cordoní, A., Caltabiano, G. & Pardo, L. Membrane protein simulations using AMBER force field and Berger lipid parameters. *J. Chem. Theory Comput.* **8**, 948–958 (2012).
203. Salomon-Ferrer, R., Case, D. A. & Walker, R. C. An overview of the Amber biomolecular simulation package. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **3**, 198–210 (2013).
204. Jakalian, A., Bush, B. L., Jack, D. B. & Bayly, C. I. Fast, efficient generation of high-quality atomic Charges. AM1-BCC model: I. Method. *J. Comput. Chem.* **21**, 132–146 (2000).
205. Jakalian, A., Jack, D. B. & Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J. Comput. Chem.* **23**, 1623–1641 (2002).

206. Duffy, E. M., Severance, D. L. & Jorgensen, W. L. Urea: potential functions, log P, and free energy of hydration. *Isr. J. Chem.* **33**, 323–330 (1993).
207. Hub, J. S. & de Groot, B. L. Mechanism of selectivity in aquaporins and aquaglyceroporins. *Proc. Natl. Acad. Sci.* **105**, 1198–1203 (2008).
208. Caleman, C., van Maaren, P. J., Hong, M., Hub, J. S., Costa, L. T. & van der Spoel, D. Force field benchmark of organic liquids: density, enthalpy of vaporization, heat capacities, surface tension, isothermal compressibility, volumetric expansion coefficient, and dielectric constant. *J. Chem. Theory Comput.* **8**, 61–74 (2011).
209. Van der Spoel, D., van Maaren, P. J. & Caleman, C. GROMACS molecule & liquid database. *Bioinformatics* **28**, 752–753 (2012).
210. Allen, T. W., Andersen, O. S. & Roux, B. Molecular dynamics—potential of mean force calculations as a tool for understanding ion permeation and selectivity in narrow channels. *Biophys. Chem.* **124**, 251–267 (2006).
211. Van Gunsteren, W. & Berendsen, H. A leap-frog algorithm for stochastic dynamics. *Mol. Simulat.* **1**, 173–185 (1988).
212. Hub, J. S., Winkler, F. K., Merrick, M. & de Groot, B. L. Potentials of mean force and permeabilities for carbon dioxide, ammonia, and water flux across a Rhesus protein channel and lipid membranes. *J. Am. Chem. Soc.* **132**, 13251–13263 (2010).
213. Gusarov, S., Pujari, B. S. & Kovalenko, A. Efficient treatment of solvation shells in 3D molecular theory of solvation. *J. Comput. Chem.* **33**, 1478–1494 (2012).
214. Sui, H., Han, B.-G., Lee, J. K., Walian, P. & Jap, B. K. Structural basis of water-specific transport through the AQP1 water channel. *Nature* **414**, 872–878 (2001).
215. Giambaşu, G. M., Luchko, T., Herschlag, D., York, D. M. & Case, D. A. Ion counting from explicit-solvent simulations and 3D-RISM. *Biophys. J* **106**, 883–894 (2014).
216. Lee, P.-H. & Helms, V. Identifying continuous pores in protein structures with PROPORES by computational repositioning of gating residues. *Proteins Struct. Funct. Bioinform.* **80**, 421–432 (2012).
217. Müller, E. M., Hub, J. S., Grubmüller, H. & de Groot, B. L. Is TEA an inhibitor for human Aquaporin-1? *Pflügers Arch.* **456**, 663–669 (2008).
218. Joung, I. S. & Cheatham III, T. E. Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J. Phys. Chem. B* **112**, 9020–9041 (2008).

219. Schrödinger, LLC. *The PyMOL Molecular Graphics System, Version 1.8* 2015.
220. Klauda, J. B., Venable, R. M., Freites, J. A., O'Connor, J. W., Tobias, D. J., Mondragon-Ramirez, C., Vorobyov, I., MacKerell Jr, A. D. & Pastor, R. W. Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types. *J. Phys. Chem. B* **114**, 7830–7843 (2010).
221. Harder, E., Anisimov, V. M., Vorobyov, I. V., Lopes, P. E. M., Noskov, S. Y., MacKerell, A. D. & Roux, B. Atomic level anisotropy in the electrostatic modeling of lone pairs for a polarizable force field based on the classical drude oscillator. *J. Chem. Theory Comput.* **2**, 1587–1597 (2006).
222. Lamoureux, G., Harder, E., Vorobyov, I. V., Roux, B. & MacKerell, A. D. A polarizable model of water for molecular dynamics simulations of biomolecules. *Chem. Phys. Lett.* **418**, 245–249 (2006).
223. Yu, H., Whitfield, T. W., Harder, E., Lamoureux, G., Vorobyov, I., Anisimov, V. M., MacKerell, A. D. & Roux, B. Simulating Monovalent and Divalent Ions in Aqueous Solution Using a Drude Polarizable Force Field. *J. Chem. Theory Comput.* **6**, 774–786 (2010).
224. Chowdhary, J., Harder, E., Lopes, P. E. M., Huang, L., MacKerell, A. D. & Roux, B. A Polarizable Force Field of Dipalmitoylphosphatidylcholine Based on the Classical Drude Model for Molecular Dynamics Simulations of Lipids. *J. Phys. Chem. B* **117**, 9142–9160 (2013).
225. Vermaas, J. V., Hardy, D. J., Stone, J. E., Tajkhorshid, E. & Kohlmeyer, A. TopoGromacs: Automated Topology Conversion from CHARMM to GROMACS within VMD. *J. Chem. Inf. Model.* **56**, 1112–1116 (2016).
226. Van Maaren, P. J. & Van Der Spoel, D. Molecular dynamics simulations of water with novel shell-model potentials. *J. Phys. Chem. B* **105**, 2618–2626 (2001).
227. Lemkul, J. A., Roux, B., Van Der Spoel, D. & Mackerell, A. D. Implementation of extended Lagrangian dynamics in GROMACS for polarizable simulations using the classical Drude oscillator model. *J. Comput. Chem.* **36**, 1473–1479 (2015).
228. Nosé, S. A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.* **81**, 511–519 (1984).
229. Hoover, W. G. Canonical dynamics: equilibrium phase-space distributions. *Phys. Rev. A: At. Mol. Opt. Phys.* **31**, 1695 (1985).

230. Vrbka, L., Vondrášek, J., Jagoda-Cwiklik, B., Vácha, R. & Jungwirth, P. Quantification and rationalization of the higher affinity of sodium over potassium to protein surfaces. *Proc. Natl. Acad. Sci.* **103**, 15440–15444 (2006).
231. Jungwirth, P. & Winter, B. Ions at aqueous interfaces: from water surface to hydrated proteins. *Annu. Rev. Phys. Chem.* **59**, 343–366 (2008).
232. Luo, Y. & Roux, B. Simulation of osmotic pressure in concentrated aqueous salt solutions. *J. Phys. Chem. Lett.* **1**, 183–189 (2009).
233. Yoo, J. & Aksimentiev, A. Improved parametrization of Li<sup>+</sup>, Na<sup>+</sup>, K<sup>+</sup>, and Mg<sup>2+</sup> ions for all-atom molecular dynamics simulations of nucleic acid systems. *J. Phys. Chem. Lett.* **3**, 45–50 (2011).
234. Rembert, K. B., Paterová, J., Heyda, J., Hilty, C., Jungwirth, P. & Cremer, P. S. Molecular mechanisms of ion-specific effects on proteins. *J. Am. Chem. Soc.* **134**, 10039–10046 (2012).
235. Hermans, J. & Wang, L. Inclusion of loss of translational and rotational freedom in theoretical estimates of free energies of binding. Application to a complex of benzene and mutant T4 lysozyme. *J. Am. Chem. Soc.* **119**, 2707–2714 (1997).
236. Hodgman, C. D., Frankel, M., *et al.* *Handbook of chemistry and physics* (Chemical Rubber Pub. Co., 1960).
237. Hub, J. S. & Awasthi, N. Probing a continuous polar defect: A reaction coordinate for pore formation in lipid membranes. *J. Chem. Theory Comput.* (2017).
238. Becke, A. Becke 3 term correlation functional (part of b3lyp). *J. Chem. Phys.* **98**, 5648 (1993).
239. Atkins, P. W. & Friedman, R. S. *Molecular quantum mechanics* (Oxford university press, 2011).
240. Neese, F. The ORCA program system. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2**, 73–78 (2012).
241. Weigend, F. & Ahlrichs, R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.* **7**, 3297–3305 (2005).
242. Zheng, J., Xu, X. & Truhlar, D. G. Minimally augmented Karlsruhe basis sets. *Theor. Chem. Acc.* **128**, 295–305 (2011).
243. Dunning Jr, T. H. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.* **90**, 1007–1023 (1989).

244. Azouzi, S., Gueroult, M., Ripoché, P., Genetet, S., Aronovicz, Y. C., Le Van Kim, C., Etchebest, C. & Mouro-Chanteloup, I. Energetic and molecular water permeation mechanisms of the human red blood cell urea transporter B. *PLoS One* **8**, e82338 (2013).
245. Bankir, L., Bouby, N., Trinh-Trang-Tan, M.-M., Ahloulay, M. & Promeneur, D. Direct and indirect cost of urea excretion. *Kidney Int.* **49**, 1598–1607 (1996).
246. Hub, J. S. & de Groot, B. L. Does CO<sub>2</sub> permeate through aquaporin-1? *Biophys. J.* **91**, 842–848 (2006).
247. Hashido, M., Kidera, A. & Ikeguchi, M. Water transport in aquaporins: osmotic permeability matrix analysis of molecular dynamics simulations. *Biophys. J.* **93**, 373–385 (2007).
248. De Groot, B. L. & Grubmüller, H. Water permeation across biological membranes: mechanism and dynamics of aquaporin-1 and GlpF. *Science* **294**, 2353–2357 (2001).
249. Tajkhorshid, E., Nollert, P., Jensen, M. Ø., Miercke, L. J., O’Connell, J., Stroud, R. M. & Schulten, K. Control of the selectivity of the aquaporin water channel family by global orientational tuning. *Science* **296**, 525–530 (2002).
250. Beitz, E., Wu, B., Holm, L. M., Schultz, J. E. & Zeuthen, T. Point mutations in the aromatic/arginine region in aquaporin 1 allow passage of urea, glycerol, ammonia, and protons. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 269–274 (2006).
251. Lamoureux, G. & Roux, B. Absolute hydration free energy scale for alkali and halide ions established from simulations with a polarizable force field. *J. Phys. Chem. B* **110**, 3308–3322 (2006).
252. Stumpe, M. C., Blinov, N., Wishart, D., Kovalenko, A. & Pande, V. S. Calculation of local water densities in biological systems: a comparison of molecular dynamics simulations and the 3D-RISM-KH molecular theory of solvation. *J. Phys. Chem. B* **115**, 319–28 (2011).
253. Kovalenko, A., Kobryn, A. E., Gusarov, S., Lyubimova, O., Liu, X., Blinov, N. & Yoshida, M. Molecular theory of solvation for supramolecules and soft matter structures: application to ligand binding, ion channels, and oligomeric polyelectrolyte gels. *Soft Matter* **8**, 1508 (2012).
254. Nikolic, D., Blinov, N., Wishart, D. & Kovalenko, A. 3D-RISM-Dock: A new fragment-based drug design protocol. *J. Chem. Theory Comput.* **8**, 3356–3372 (2012).
255. Imai, T., Hiraoka, R., Kovalenko, A. & Hirata, F. Water molecules in a protein cavity detected by a statistical-mechanical theory. *J. Am. Chem. Soc.* **127**, 15334–15335 (2005).



256. Imai, T, Kovalenko, A & Hirata, F. Hydration structure, thermodynamics, and functions of protein studied by the 3D-RISM theory. *Molec. Sim.* **32**, 817–824 (2006).
257. Fusani, L., Wall, I., Palmer, D. & Cortes, A. Optimal water networks in protein cavities with GASol and 3D-RISM. *Bioinformatics* (2018).
258. Wang, Z., Yu, T., Sang, J.-P., Zou, X.-W., Yan, C. & Zou, X. Computation and Simulation of the Structural Characteristics of the Kidney Urea Transporter and Behaviors of Urea Transport. *J. Phys. Chem. B* **119**, 5124–5131 (2015).
259. Cao, S., Sheong, F. K. & Huang, X. Reference interaction site model with hydrophobicity induced density inhomogeneity: An analytical theory to compute solvation properties of large hydrophobic solutes in the mixture of polyatomic solvent molecules. *J. Chem. Phys.* **143**, 054110 (2015).
260. Kast, S. M. & Kloss, T. Closed-form expressions of the chemical potential for integral equation closures with certain bridge functions. *J. Chem. Phys.* **129**, 236101 (2008).
261. Hub, J. S. & de Groot, B. L. Comment on “Molecular selectivity in aquaporin channels studied by the 3D-RISM theory”. *J. Phys. Chem. B* **115**, 8364–6; 8367–9 (2011).
262. Phongphanphanee, S., Yoshida, N. & Hirata, F. The potential of mean force of water and ions in aquaporin channels investigated by the 3D-RISM method. *J. Mol. Liq.* **147**, 107–111 (2009).
263. Kleywegt, G. J. Validation of protein crystal structures. *Acta Crystallogr.* **D56**, 249–265 (2000).
264. Kleywegt, G. J. Crystallographic refinement of ligand complexes. *Acta Crystallogr.* **D63**, 94–100 (2007).
265. Cooper, D. R., Porebski, P. J., Chruszcz, M. & Minor, W. X-ray crystallography: Assessment and validation of protein-small molecule complexes for drug discovery. *Expert Opin. Drug Discov.* **6**, 771–782 (2011).
266. Doyle, D. A., Cabral, J. M., Pfuetzner, R. A., Kuo, A., Gulbis, J. M., Cohen, S. L., Chait, B. T. & MacKinnon, R. The structure of the potassium channel: molecular basis of K<sup>+</sup> conduction and selectivity. *Science* **280**, 69–77 (1998).
267. Zhou, Y. & MacKinnon, R. The occupancy of ions in the K<sup>+</sup> selectivity filter: charge balance and coupling of ion binding to a protein conformational change underlie high conduction rates. *J. Mol. Biol.* **333**, 965–975 (2003).

268. Kratochvil, H. T., Carr, J. K., Matulef, K., Annen, A. W., Li, H., Maj, M., Ostmeyer, J., Serrano, A. L., Raghuraman, H, Moran, S. D., *et al.* Instantaneous ion configurations in the K<sup>+</sup> ion channel selectivity filter revealed by 2D IR spectroscopy. *Science* **353**, 1040–1044 (2016).
269. Nam, K. H., Sung, M. W. & Hwang, K. Y. Structural insights into the substrate recognition properties of  $\beta$ -glucosidase. *Biochem. Biophys. Res. Comm.* **391**, 1131–1135 (2010).
270. Botos, I., Wu, Z., Lu, W. & Wlodawer, A. Crystal structure of a cyclic form of bovine pancreatic trypsin inhibitor. *FEBS Lett.* **509**, 90–94 (2001).
271. Morozenko, A & Stuchebrukhov, A. Dowser++, a new method of hydrating protein structures. *Proteins: Struct. Func. Bioinf.* **84**, 1347–1357 (2016).
272. Sridhar, A., Ross, G. A. & Biggin, P. C. Waterdock 2.0: Water placement prediction for Holo-structures with a pymol plugin. *PLoS ONE* **12**, e0172743 (2017).
273. Hub, J. S., Groot, B. L. D., Grubmueller, H. & Groenhof, G. Quantifying artifacts in Ewald simulations of inhomogeneous systems with a net charge. *J. Chem. Theory Comput.* **10** (2014).
274. Figueirido, F., Del Buono, G. S. & Levy, R. M. On finite-size effects in computer simulations using the Ewald potential. *J. Chem. Phys.* **103**, 6133–6142 (1995).
275. Figueirido, F., Del Buono, G. S. & Levy, R. M. On finite-size corrections to the free energy of ionic hydration. *J. Phys. Chem. B* **101**, 5622–5623 (1997).
276. Hummer, G., Pratt, L. R. & García, A. E. Ion sizes and finite-size corrections for ionic-solvation free energies. *J. Chem. Phys.* **107**, 9275–9277 (1997).
277. Hummer, G., Pratt, L. R. & García, A. E. Molecular theories and simulation of ions and polar molecules in water. *J. Phys. Chem. A* **102**, 7885–7895 (1998).
278. Sakane, S., Ashbaugh, H. S. & Wood, R. H. Continuum corrections to the polarization and thermodynamic properties of Ewald sum simulations for ions and ion pairs at infinite dilution. *J. Phys. Chem. B* **102**, 5673–5682 (1998).
279. Hünenberger, P. H. & McCammon, J. A. Ewald artifacts in computer simulations of ionic solvation and ion–ion interaction: a continuum electrostatics study. *J. Chem. Phys.* **110**, 1856–1872 (1999).

280. Hünenberger, P. H. & McCammon, J. A. Effect of artificial periodicity in simulations of biomolecules under Ewald boundary conditions: a continuum electrostatics study. *Biophys. Chem.* **78**, 69–88 (1999).
281. Kastenholz, M. A. & Hünenberger, P. H. Influence of artificial periodicity and ionic strength in molecular dynamics simulations of charged biomolecules employing lattice-sum methods. *J. Phys. Chem. B* **108**, 774–788 (2004).
282. Hoshi, T., Zagotta, W. N. & Aldrich, R. W. Shaker potassium channel gating. I: Transitions near the open state. *J. Gen. Physiol.* **103**, 249–278 (1994).
283. Kuo, A., Gulbis, J. M., Antcliff, J. F., Rahman, T., Lowe, E. D., Zimmer, J., Cuthbertson, J., Ashcroft, F. M., Ezaki, T. & Doyle, D. A. Crystal structure of the potassium channel KirBac1.1 in the closed state. *Science* **300**, 1922–1926 (2003).
284. Hilf, R. J. & Dutzler, R. Structure of a potentially open state of a proton-activated pentameric ligand-gated ion channel. *Nature* **457**, 115 (2009).
285. Heer, F. T., Posson, D. J., Wojtas-Niziurski, W., Nimigean, C. M. & Berneche, S. Mechanism of activation at the selectivity filter of the KcsA K<sup>+</sup> channel. *eLife* **6** (2017).
286. McIlwain, B. C., Newstead, S. & Stockbridge, R. B. Cork-in-Bottle Occlusion of Fluoride Ion Channels by Crystallization Chaperones. *Structure* **26**, 635–639 (2018).
287. Saparov, S. M. & Pohl, P. Beyond the diffusion limit: Water flow through the empty bacterial potassium channel. *Proc. Nat. Acad. Sci.* **101**, 4805–4809 (2004).
288. Pohl, P. Combined transport of water and ions through membrane channels. *Biol. Chem.* **385**, 921–926 (2004).
289. Furini, S., Beckstein, O. & Domene, C. Permeation of water through the KcsA K<sup>+</sup> channel. *Proteins: Struct., Funct., Bioinform.* **74**, 437–448 (2009).
290. Ulmschneider, M. B., Bagnéris, C., McCusker, E. C., DeCaen, P. G., Delling, M., Clapham, D. E., Ulmschneider, J. P. & Wallace, B. A. Molecular dynamics of ion transport through the open conformation of a bacterial voltage-gated sodium channel. *Proc. Nat. Acad. Sci.* **110**, 6364–6369 (2013).
291. Chakravarty, S., Ung, A. R., Moore, B., Shore, J. & Alshamrani, M. A Comprehensive Analysis of Anion–Quadrupole Interactions in Protein Structures. *Biochemistry* **57**, 1852–1867 (2018).
292. Sanders, C. R. & Sönnichsen, F. Solution NMR of membrane proteins: practice and challenges. *Magn. Reson. Chem.* **44** (2006).

293. Opella, S. J. & Marassi, F. M. Applications of NMR to membrane proteins. *Arch. Biochem. Biophys.* **628**, 92–101 (2017).
294. Zhao, X. in *NMR of Proteins and Small Biomolecules* 187–213 (Springer, 2011).
295. Visscher, K., Medeiros-Silva, J., Mance, D., Rodrigues, J. P., Daniëls, M., Bonvin, A. M., Baldus, M. & Weingarh, M. Supramolecular organization and functional implications of K<sup>+</sup> channel clusters in membranes. *Angew. Chem.* (2017).
296. Hwang, H., Schatz, G. C. & Ratner, M. A. Steered molecular dynamics studies of the potential of mean force of a Na<sup>+</sup> or K<sup>+</sup> ion in a cyclic peptide nanotube. *J. Phys. Chem. B* **110**, 26448–26460 (2006).
297. Patargias, G., Martay, H. & Fischer, W. B. Reconstructing potentials of mean force from short steered molecular dynamics simulations of Vpu from HIV-1. *J. Biomol. Struct. Dyn.* **27**, 1–12 (2009).
298. Sauguet, L., Poitevin, F., Murail, S., Van Renterghem, C., Moraga-Cid, G., Malherbe, L., Thompson, A. W., Koehl, P., Corringer, P.-J., Baaden, M., *et al.* Structural basis for ion permeation mechanism in pentameric ligand-gated ion channels. *EMBO J.* **32**, 728–741 (2013).
299. Paz, S. A., Maragliano, L. & Abrams, C. F. Effect of Intercalated Water on Potassium Ion Transport through Kv1. 2 Channels Studied via On-the-Fly Free-Energy Parametrization. *J. Chem. Theory Comput.* **14**, 2743–2750 (2018).
300. Yeromin, A. V., Zhang, S. L., Jiang, W., Yu, Y., Safrina, O. & Cahalan, M. D. Molecular identification of the CRAC channel by altered ion selectivity in a mutant of Orai. *Nature* **443**, 226 (2006).
301. McKinnon, N. K., Reeves, D. C. & Akabas, M. H. 5-HT<sub>3</sub> receptor ion size selectivity is a property of the transmembrane channel, not the cytoplasmic vestibule portals. *J. Gen. Physiol.* **138**, 453–466 (2011).
302. Finkelstein, A. & Cass, A. Effect of cholesterol on the water permeability of thin lipid membranes. *Nature* **216**, 717 (1967).
303. Rubenstein, J., Smith, B. A. & McConnell, H. M. Lateral diffusion in binary mixtures of cholesterol and phosphatidylcholines. *Proc. Nat. Acad. Sci.* **76**, 15–18 (1979).
304. Miao, L., Nielsen, M., Thewalt, J., Ipsen, J. H., Bloom, M., Zuckermann, M. J. & Mouritsen, O. G. From lanosterol to cholesterol: structural evolution and differential effects on lipid bilayers. *Biophys. J.* **82**, 1429–1444 (2002).

305. Karatekin, E., Sandre, O., Guitouni, H., Borghi, N., Puech, P.-H. & Brochard-Wyart, F. Cascades of transient pores in giant vesicles: line tension and transport. *Biophys. J.* **84**, 1734–1749 (2003).
306. Róg, T., Pasenkiewicz-Gierula, M., Vattulainen, I. & Karttunen, M. Ordering effects of cholesterol and its analogues. *Biochim. Biophys. Acta, Biomembr.* **1788**, 97–121 (2009).
307. Veatch, S. L. & Keller, S. L. Organization in lipid membranes containing cholesterol. *Phys. Rev. Lett.* **89**, 268101 (2002).
308. Tsamaloukas, A., Szadkowska, H. & Heerklotz, H. Thermodynamic comparison of the interactions of cholesterol with unsaturated phospholipid and sphingomyelins. *Biophys. J.* **90**, 4479–4487 (2006).
309. Ali, M. R., Cheng, K. H. & Huang, J. Assess the nature of cholesterol–lipid interactions through the chemical potential of cholesterol in phosphatidylcholine bilayers. *Proc. Nat. Acad. Sci.* **104**, 5372–5377 (2007).
310. Pandit, S. A., Chiu, S.-W., Jakobsson, E., Grama, A. & Scott, H. Cholesterol packing around lipids with saturated and unsaturated chains: a simulation study. *Langmuir* **24**, 6858–6865 (2008).
311. Almeida, P. F. Thermodynamics of lipid interactions in complex bilayers. *Biochim. Biophys. Acta, Biomembr.* **1788**, 72–85 (2009).
312. Heerklotz, H. & Tsamaloukas, A. Gradual change or phase transition: characterizing fluid lipid-cholesterol membranes on the basis of thermal volume changes. *Biophys. J.* **91**, 600–607 (2006).
313. Marsh, D. Liquid-ordered phases induced by cholesterol: a compendium of binary phase diagrams. *Biochim. Biophys. Acta, Biomembr.* **1798**, 688–699 (2010).
314. Waheed, Q., Tjörnhammar, R. & Edholm, O. Phase transitions in coarse-grained lipid bilayers containing cholesterol by molecular dynamics simulations. *Biophys. J.* **103**, 2125–2133 (2012).
315. Toppozini, L., Meinhardt, S., Armstrong, C. L., Yamani, Z., Kučerka, N., Schmid, F. & Rheinstädter, M. C. Structure of cholesterol in lipid rafts. *Phys. Rev. Lett.* **113**, 228101 (2014).
316. Díaz-Tejada, C., Ariz-Extreme, I., Awasthi, N. & Hub, J. S. Quantifying lateral inhomogeneity of cholesterol-containing membranes. *J. Phys. Chem. Lett.* **6**, 4799–4803 (2015).
317. Risselada, H. J. & Marrink, S. J. The molecular face of lipid rafts in model membranes. *Proc. Nat. Acad. Sci.* **105**, 17367–17372 (2008).

318. Schäfer, L. V., de Jong, D. H., Holt, A., Rzepiela, A. J., de Vries, A. H., Poolman, B., Killian, J. A. & Marrink, S. J. Lipid packing drives the segregation of transmembrane helices into disordered lipid domains in model membranes. *Proc. Nat. Acad. Sci.* **108**, 1343–1348 (2011).
319. Tolpekina, T., den Otter, W. K. & Briels, W. J. Nucleation free energy of pore formation in an amphiphilic bilayer studied by molecular dynamics simulations. *J. Chem. Phys.* **121**, 12060–12066 (2004).
320. Halling, K. K., Ramstedt, B., Nyström, J. H., Slotte, J. P. & Nyholm, T. K. Cholesterol interactions with fluid-phase phospholipids: effect on the lateral organization of the bilayer. *Biophys. J.* **95**, 3861–3871 (2008).
321. Pan, J., Mills, T. T., Tristram-Nagle, S. & Nagle, J. F. Cholesterol perturbs lipid bilayers nonuniversally. *Phys. Rev. Lett.* **100**, 198103 (2008).
322. Marrink, S. J., de Vries, A. H., Harroun, T. A., Katsaras, J. & Wassall, S. R. Cholesterol shows preference for the interior of polyunsaturated lipid membranes. *J. Am. Chem. Soc.* **130**, 10–11 (2008).
323. Leventis, R. & Silvius, J. R. Use of cyclodextrins to monitor transbilayer movement and differential lipid affinities of cholesterol. *Biophys. J.* **81**, 2257–2267 (2001).
324. Niu, S.-L. & Litman, B. J. Determination of membrane cholesterol partition coefficient using a lipid vesicle–cyclodextrin binary system: effect of phospholipid acyl chain unsaturation and headgroup composition. *Biophys. J.* **83**, 3408–3415 (2002).

---

## CURRICULUM VITAE

---

IGOR ARIZ EXTREME

### *Personal Information*

DATE AND PLACE OF BIRTH: 25<sup>th</sup> of February, 1988 in Donostia

ADDRESS: Ewaldstr. 25, 37085 Göttingen

E-MAIL: iarizex@pm.me

### *Education*

2007-2012 **Bachelor in Biochemistry and Molecular Biology**

University of Basque Country, Bilbo, Spain

B. Sc. Thesis: *Electrophysiology and Model Cells*

2012-2014 **Master in Bioinformatics for Health Sciences**

Pompeu Fabra University, Barcelona, Spain

M. Sc. Thesis: *Deep unsupervised learning for MD simulations*

2014-2018 **Ph. D. student at Computational Molecular Biophysics Group**

Georg-August-Universität, Göttingen, Germany

Ph. D. Thesis: *Computational Studies of Small Molecule Permeation across Membrane Channels*

### *Publications*

Ariz-Extreme, I. & Hub, J. S. Assigning crystallographic electron densities with free energy calculations – The case of the fluoride channel Fluc. *PLoS one* **13**, e0196751 (2018).

Ariz-Extreme, I. & Hub, J. S. Potential of mean force calculations of solute permeation across UT-B and AQP1: A comparison between molecular dynamics and 3D-RISM. *J. Phys. Chem. B* **121**, 1506–1519 (2017).

Doerr, S., Ariz-Extreme, I., Harvey, M. J. & De Fabritiis, G. Dimensionality reduction methods for molecular simulations. *arXiv preprint arXiv:1710.10629* (2017).

Díaz-Tejada, C., Ariz-Extreme, I., Awasthi, N. & Hub, J. S. Quantifying lateral inhomogeneity of cholesterol-containing membranes. *J. Phys. Chem. Lett.* **6**, 4799-4803 (2015).