

Understanding the molecular machinery of
aquaporins through molecular dynamics
simulations.

Dissertation
for the award of the degree
"Doctor rerum naturalium"
Division of Mathematics and Natural Sciences
of the Georg-August-Universität Göttingen

submitted by
Camilo Andrés Aponte-Santamaría
from Bogotá (Colombia)

Göttingen, 2011

Thesis committee members:

- Prof. Dr. Bert de Groot (supervisor)
Computational and Biomolecular Dynamics Group,
Max Planck Institute for Biophysical Chemistry

- Prof. Dr. Christian Griesinger (reviewer)
Department of NMR-based Structural Biology,
Max Planck Institute for Biophysical Chemistry

- Prof. Dr. Marcus Müller (reviewer)
Institute for Theoretical Physics,
Georg August University Göttingen

It is declared that the present thesis has been written independently and with no other sources and aids than quoted.

Camilo Aponte-Santamaría

Abstract

Aquaporins are protein channels responsible for the permeation of water and other small solutes through biological membranes in response to osmotic pressure. The goal of the present thesis is to expand our understanding on the molecular machinery of aquaporins by employing molecular dynamics simulations and related computational techniques.

First, we provide a solute permeation mechanism for the solute permeation through the *Plasmodium falciparum* aquaglyceroporin that is a promising antimalarial drug target. In this mechanism, hydrophobic regions in the middle of the channel are the main water rate limiting barriers. In addition, the replacement of water-arginine interactions and solute-matching at the most constricted region of the channel are the main determinants underlying selectivity for the permeation of solutes like glycerol and urea.

Second, we investigate the molecular determinants governing aquaporin gating, which has emerged as an efficient regulatory mechanism for organisms to quickly counteract sudden osmotic shocks. Our simulations, together with structural and functional studies, suggest that the yeast aquaporin-1 may be gated by both phosphorylation of a serine residue or mechanosensing. In addition, for spinach plant aquaporin, our simulations confirm that cytosolic loop D provides the necessary machinery to trigger the channel gating. However, they do not support a gating mechanism mediated serine-phosphorylation or histidine-protonation, as proposed in the current gating model. Furthermore, we observed voltage regulation of the single-channel water permeability of human AQP4 *in silico*, attributed to gating transitions of the arginine residue at the aromatic/arginine region. Our results, together with similar observations made for human AQP1, suggests that voltage sensitivity may be a general feature of AQPs, a hypothesis to be tested experimentally.

Third, we study three important processes where aquaporins interact with different types of (bio)molecules. The formation and stability of the AQP2-LIP5 complex, which is crucial for the trafficking of AQP2 in renal cells, was investigated. Our results predict at least two putative structures of the complex, with the aquaporin tetramer being embedded in a membrane, stabilized by the binding of three leucines at the C-terminus in AQP2 to a hydrophobic cleft in LIP5. Moreover, we observed that AQP0-lipid interactions mainly depend on the matching of the lipid tails to the AQP0 surface, rather than their headgroups, and that lipids gradually adopt bulk-lipid properties when located further away from AQP0. Our data also supports that specific lipid positions, that had been observed in a two-dimensional crystal of AQP0 by electron crystallography, are indeed representative of those adopted by lipids around AQP0 tetramers immersed at low concentrations in a lipid bilayer. Finally, we describe how molecular dynamics simulations can be used, in combination with experimental functional assays and molecular docking calculations, in the search and refinement of putative AQP9 blockers.

Acknowledgements

I initially would like to thank Prof. Bert de Groot for giving me the opportunity to do my doctoral studies in his group. His excellent supervision and constant support has been crucial to achieve the goals I addressed in this project. He always has the disposition and patience to listen to me, giving wise advice and suggestions on how to solve problems and doubts, and encouraging me to address science from different creative perspectives.

I thank Prof. Helmut Grubmüller who established the department of theoretical and computational biophysics, gathering the best team of people and infrastructure to do research in this field. I am very grateful to all the present and former members of the department. I have learned and enjoyed a lot, doing science surrounded by them. I particularly thank Eveline Heinemann for making our life easier, taking care of all the details of the daily office life. Also, many thanks to Ansgar Esztermann and Martin Fechner for maintaining the cluster and the workstations working perfectly. I would also like to thank Carsten Kutzner for sharing his programming experience and helping with the use of the GROMACS package, and Ulrike Gerischer for carefully reading some of the manuscripts I wrote, including the present thesis. Thanks to Christian Blau, Vytantas Gapsys, Gerrit Groenhof, Martin Höfling, David Köpfer, Dirk Matthes, Chen Song, Martin Vesper, Esteban Vöhringer, Sören Wacker, Maarten Wolf and Ulrich Zachariae. I specially want to acknowledge Jochen Hub for the helpful discussions and explanations concerning aquaporins and physics, and Daniel Seeliger for the help with all the practical problems and programming.

I thank Prof. Marcus Müller and Prof. Christian Griesinger, the other two members of my thesis committee, for finding the time to meet and discuss about the progress of my project. I would also like to acknowledge Antje Erdmann, Kirsten Pöhlker and Michaela Böttcher for the assistance and guidance with all the aspects concerning the Göttingen Graduate School for Neurosciences and Molecular Biosciences (GGNB).

This thesis was supported by grants from the European Commission, the Marie Curie Research Training Network of Aqua(glycero)porins (MRTNCT-2006-035995). I am extremely grateful to all the people involved in this multidisciplinary network for all the collaborations, and the exchange of information and ideas. Among them are: Gerhard Fischer, Urszula Kosinska-Eriksson, Stephan Hohmann, Richard Neutze and Karin Lindkvist-Petersson, at the Göteborg University (Göteborg, Sweden); Andreas Engel and W. Kukulski, at Biozentrum (Basel, Switzerland); Peter Deen, Hanka Venselaar and Sanne Nabuurs, at the Nijmegen Medical Center (Nijmegen, Netherlands), and Michael Rützler, at the University of Aarhus (Aarhus, Denmark). Many thanks also to Tom Walz and Richard Hite, at the Harvard University (USA), for helpful discussions.

Outside the laboratory, I would like to thank many friends, who have made my life in Göttingen more enjoyably and "climbable". Thanks to Nicole, who has accompanied me in very difficult situations and invited me to join the climbing group; Guillem and Wilson for many great and relaxing moments; Adriana and Eduardo, my godparents here in Göttingen, and Nadine for giving me the opportunity to share part of our lives together. Thanks to the other members of the climbing group: Birgit, Frank, Rado, Suzannah, Yana and Yoshi. Thanks to my flatmates Annalisa, Jan Valentin, Mateusz and Yoshi for making our WG a home-like place. To all the other friends from the Spanish-spoken community, particularly Abud, Laura, Matias, Natalia and Rodolfo, thank you very much.

On the other side of the Atlantic ocean, in Colombia, i want to express infinite gratitude to my parents, Blanca and Agustín; my brother, Diego, and my other relatives. They are the source of inspiration of all the things that happen in my life and the reason to continue further. I also thank all my school and university friends: Angélica, David, Daniel, Dilian, Gabriel, Jaime, Jorge, Karen, José Daniel, Juliana, Lola, Marcela, Mónica, Natalia, Veronica, Viviana and William.

Finally, I dedicate the present thesis to Frenly Marcela. She was always very happy and enthusiastic about all my scientific achievements. Her harmony, happiness and beauty will always enlighten my way. Her memory is and is going to be present in my mind and heart forever.

Contents

1	Introduction	1
2	Theory and Methods	11
2.1	Molecular dynamics simulations	11
2.1.1	Approximations of MD simulations	12
2.1.2	Numerical solution of the equations of motion	14
2.1.3	Simulation details	14
2.2	Observables derived from MD simulations	16
2.2.1	Water permeability coefficients	16
2.2.2	Potentials of mean force	17
2.2.3	Principal component analysis	22
3	Solute permeation through the <i>Plasmodium falciparum</i> aquaglyceroporin	25
3.1	Introduction	25
3.2	Theory and methods	26
3.2.1	Equilibrium molecular dynamics simulations	26
3.2.2	Water permeability calculations	28
3.2.3	Pore dimensions	28
3.2.4	Potential of mean force for water	28
3.2.5	Umbrella sampling simulations	29
3.3	Results	30
3.4	Discussion	33
3.4.1	Quantifying the water transport activity	33

3.4.2	The energetics of water permeation	34
3.4.3	The effect of point mutations	35
3.4.4	Permeation of other solutes	36
3.5	Conclusions	37
4	The gating mechanism of aquaporins	39
4.1	The yeast aquaporin: gating mediated by phosphorylation and mechanosensing	40
4.1.1	Introduction	40
4.1.2	Methods	40
4.1.3	Results	43
4.1.4	Discussion and conclusion	45
4.2	Spinach aquaporin gating: phosphorylation or histidine-protonation?	49
4.2.1	Introduction	49
4.2.2	Methods	50
4.2.3	Results	53
4.2.4	Discussion	56
4.3	Is aquaporin 4 a voltage-regulated channel?	58
4.3.1	Introduction	58
4.3.2	Methods	59
4.3.3	Results and Discussion	62
4.3.4	Conclusion	64
5	Interaction of aquaporins with other (bio)molecules	65
5.1	the complex formed by aquaporin 2 and the LIP5 protein	66
5.1.1	Introduction	66
5.1.2	Methods	67
5.1.3	Results and Discussion	70
5.1.4	Conclusion and outlook	75
5.2	Aquaporin 0: an excellent model to study protein-lipid interactions	77
5.2.1	Introduction	77
5.2.2	Methods	77
5.2.3	Results and discussion	79
5.2.4	Summary and outlook	82
5.3	Inhibition of aquaporin 9	85
5.3.1	Introduction	85
5.3.2	Methods	85

5.3.3 Preliminary results	88
5.3.4 Summary and outlook	90
6 Summary and conclusions	93
bibliography	115
CURRICULUM VITAE: Camilo Andrés Aponte-Santamaría	

CHAPTER 1

Introduction

Aquaporins (AQPs) are protein channels inserted in biological membranes to facilitate the permeation of water and other small neutral solutes in response to osmotic pressure [1].

Why do all living organisms need aquaporins? Water is essential for life. Cells are constituted in a great proportion by water. Water is also the environment where cells are immersed in and where most of the intracellular reactions take place [2]. Cells need to keep tight control of the water content, and thereby maintain the water homeostasis, by allowing the rapid uptake or expulsion of water. Since biological membranes are almost impermeable by water, AQPs have evolved to perform this fundamental homeostatic task. These proteins allow the passage of water through the cell membrane in response to osmotic pressure, caused by solute concentration differences between the intracellular and the extracellular media [3]. Remarkably, these proteins carry out their function in an extremely efficient manner, with several thousands of millions ($\sim 10^9$) of water molecules spontaneously permeating the channel per second by diffusion [4].

In addition to water, AQPs facilitate the entrance of other small solutes necessary for the cell (such as glycerol) or the release of waste products (like urea or ammonia) [5]. However, they strictly exclude the permeation of charged molecules, in particular protons, in order to maintain vital electrochemical gradients across the cell membrane.

AQPs have been subject of intense research over the last two decades. A large number of experimental and computational studies have contributed to explain how aquaporins perform their function, how they are implicated in crucial physiological processes in different living

organisms, and how several diseases are associated with their malfunction. In the following sections, the state of the art in the aquaporin field and the goals of the present thesis are briefly discussed.

The aquaporin family

AQPs can be classified into two major groups: aquaporins, if they are only permeable to water, and aquaglyceroporins, if in addition to water they facilitate the permeation of other small uncharged solutes, such as glycerol, urea or ammonia [6, 7]. AQPs are present in all kingdoms of life [8]. In mammals, 13 different AQPs have been identified (AQP0-AQP12) [7]. AQP0, AQP1, AQP2, AQP4, AQP5, AQP6 and AQP8 are classified as aquaporins, whereas, AQP3, AQP7, AQP9 and AQP10 are categorized as aquaglyceroporins [6, 7]. AQP11 and AQP12, whose function is still not well understood, share only 20 % of the aminoacid sequence similarity with the other aqua(glycero)porins and may represent a third evolutionary branch within the aquaporin family [9–11]. Plants express an abundant number of aquaporin homologs [12]. Accordingly, between 35 and 38 aquaporin gene sequences have been identified for *Arabidopsis thaliana* [13, 14], 31 for maize [15] and 33 for rice [16]. Examples of AQPs present in bacteria are AqpZ and GlpF in *E. coli* [17, 18], which are an aquaporin and a aquaglyceroporin, respectively, and AqpM in *Methanothermobacter marburgensis* [19, 20], being in a unique subdivision between the two major aquaporin families [21]. In fungi, up to five aquaporin genes are encoded [22], and in several protozoan parasites a single aquaporin sequence was identified [23–29].

Quantifying the water and solute permeation through AQPs

AQPs have been characterized in terms of their capability to conduct water and other solutes. The water permeability can be quantified in a setup consisting of two compartments separated by a membrane that contains the AQP under study, by inducing osmotic pressure differences between the compartments. An example of this is the use of *Xenopus laevis* oocytes, where the water permeability is assessed by measuring cell volume changes upon hypotonic stress [1, 30]. This technique successfully led to the discovery of the AQPs [1] and has nowadays become an established method to measure water permeabilities [30]. Alternatively, other aquaporin functional assays have been used such as reconstituted proteoliposomes [31], 2D crystals [32], or suspended lipid bilayers [33]. The permeation of other solutes (i.e. glycerol or urea) can be studied in a similar two-compartment experimental setup, by creating a chemical gradient between the two compartments with equal osmolality.

Functional characterization by employing the mentioned techniques revealed that AQPs are

extremely efficient water channels. For instance, in the pioneering experiments carried out by Preston *et al.* [1], oocytes encoding AQP1 swelled up to an increase of 40% in their volume within 2 min, when transferred from a 200 mosM to a 70 mosM buffer. Accordingly, the single-channel water permeability for AQP1 was later estimated to be between 4.6 and $5.4 \times 10^{-14} \text{ cm}^3/\text{s}$ [31, 32]. In consequence, if there is a solute concentration difference of 300 mM between the intracellular and the extracellular media, then a *net* flux of about 10 million of water molecules passing through an AQP1 monomer per second is established.

AQPs display a broad range of water permeabilities. For AQP0 [34], for example, the single-channel water permeability was estimated to be approximately one order of magnitude smaller than that for AQP1. On the other side of the spectrum, AQPs such as AQP4 [34, 35] or AQPZ [18, 33] displayed single-channel water permeabilities several fold larger than the value for AQP1.

In addition to water, the permeation of other solutes was also investigated, allowing to make the classification between aquaporins and aquaglyceroporins. Furthermore, functional assays with aquaporin mutants were found extremely useful to identify residues and regions of the protein which are important for the solute permeation.

The structure of AQPs

Eighteen 3D structures of AQPs (and aquaporin mutants) with a resolution better than 5 Å have been obtained so far by X-ray crystallography or electron microscopy. For a review of the structure of AQPs we refer to Walz *et al.* [8]. These structures are bovine [36] and ovine [37–39] AQP0; human [40, 41] and bovine [42] AQP1; human [43] and rat [44, 45] AQP4; human AQP5 [46]; spinach plant aquaporin (SoPIP1;2) [47–49]; the glycerol facilitator (GlpF) [50, 51] and AQPZ [52, 53] from *Escherichia coli*; archaeal AQPM [21]; Aqy1 from the yeast *Pichia pastoris* [54], and PfAQP [55] from the *Plasmodium falciparum* malarial parasite.

AQPs form tetramers with each monomer constituting an independent pore (figure 1.1A). Each monomer is composed of six transmembrane helices (H1-H6) and two short helical segments (HB and HE) (figures 1.1B-C). The first half of the protein (H1-H3 and HB) is related to the second half (H4-H6 and HE) by a quasi two-fold symmetry in the plane of the membrane. Both short helices HB and HE have a highly conserved asparagine-proline-alanine (NPA) motif, a hallmark of AQP sequences. These two helices meet in the middle of the protein, stabilized by attractive van der Waals interactions between the NPA motifs. AQPs are fairly rigid proteins due to the tetrameric arrangement and a tight packing of the helices. The pore connecting the intracellular and the extracellular side is about 25 Å in length (figure 1.1D-E). The most constricted region of the pore is located around 7 Å from the NPA region

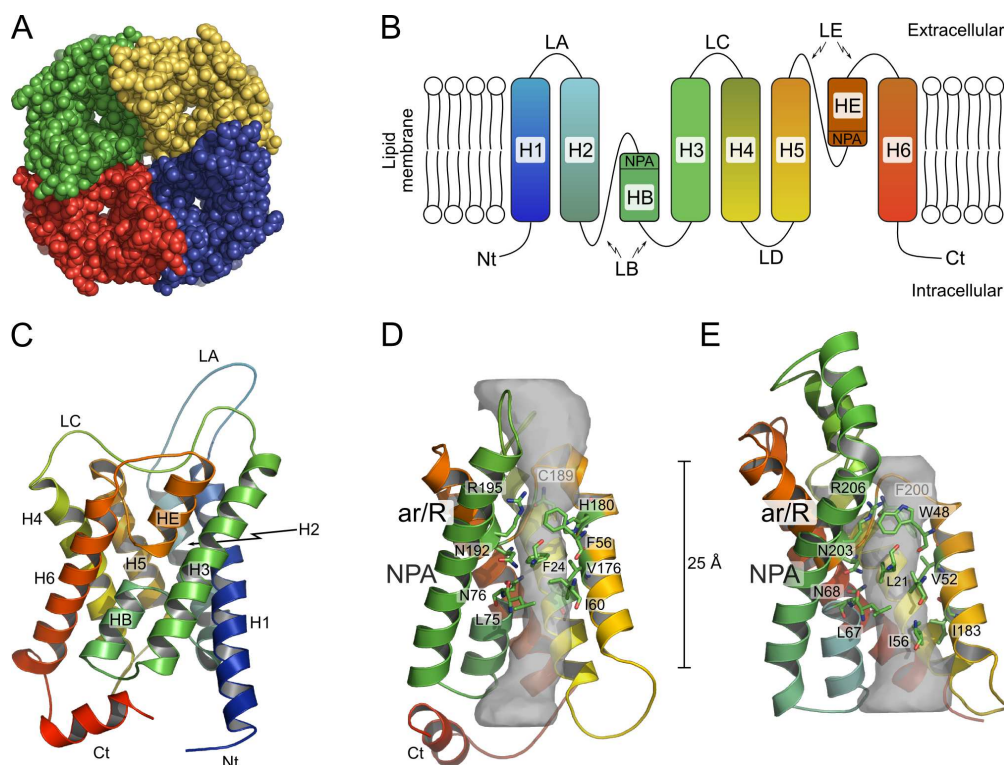


Figure 1.1. The structure of aquaglyceroporins. *A.* Top view of the tetrameric structure with each monomer (in different color) constituting an independent pore. *B.* Secondary structure of each monomer. It consists of six transmembrane helices (H1-H6) and two short helical segments (HB and HE). The highly conserved asparagine-proline-alanine (NPA) motifs at the helices HB and HE are also indicated. *C-E.* Cartoon representation of AQP1 (*C-D*) and GlpF (*E*), a water-specific aquaporin and an aquaglyceroporin, respectively. In *D* and *E*, the aromatic/arginine (ar/R) constriction site and the NPA region are highlighted. Residues facing the pore at these two constriction regions are displayed in stick representation. The pore connecting the intracellular and the extracellular media is depicted by the grey surface. It is ~ 25 Å length, being the ar/R site the narrowest point with diameters of 2.8 Å and 3.4 Å for AQP1 and GlpF, respectively.

towards the extracellular side. This region, termed the aromatic/arginine (ar/R) region, is formed by a highly conserved arginine residue juxtaposed to two aromatic residues, and constitutes another key structural motif among the AQP family. For water-specific aquaporins, the ar/R site is 2.8 Å in diameter, approximately the size of a water molecule, whereas, for aquaglyceroporins, it is 3.4 Å in diameter, matching to the size of the carbon hydroxyl group of polyols such as glycerol. The NPA region is a second constriction site, but wider than the ar/R region. There, the asparagine residues of the NPA motifs are facing the pore, together with two rings of hydrophobic residues.

The mechanism of permeation and selectivity of AQPs

How do AQPs achieve such large solute permeation rates while maintaining strict selectivity, only permitting the passage of small uncharged solutes? Molecular dynamics simulations, starting from the atomistic 3D structures, gave insights into this issue, by providing dynamical and energetical information that is usually difficult to access experimentally [56, 57].

Water permeation through AQPs occurs in a highly coordinated quasi single-file manner [4, 51]. A fine-tuned hydrogen bond network is established between the permeant water molecules and the protein residues facing the pore, which results to be complementary to the network in the aqueous media. Inside the channel, the rupture of water-water hydrogen bonds is largely compensated by protein-water interactions, reducing the energetic cost for the water permeation. The channel therefore provides an energetically favourable permeation pathway, which allows the observed high permeation rates. The NPA region was observed to be the site with the highest energy barrier for the water permeation, due to the two rings of hydrophobic residues located there. This region was thus identified as the main water rate-limiting region. The narrow geometry of the pore and its amphiphatic nature were observed to allow AQPs to maintain rigorous control of the solute permeation. The ar/R region was identified to act as the main selectivity filter [4, 58]. In water-specific aquaporins, the permeation of solutes larger than water is restricted by the size of the pore at this critical region. Furthermore, the replacement of water-arginine interactions and the solute-matching in the hydrophobic pocket at this constriction site are the two main determinants underlying the selectivity for the permeation of other solutes like glycerol and urea [58]. In addition, permeation of charged molecules, in particular protons, through AQPs was found to be prevented by an electrostatic barrier at the NPA region, created by the macrodipoles of the short helices HB and HE [59–61]. Moreover, AQPs were suggested to be permeable to CO₂ only if they are embedded in lipid membranes with unusually low CO₂ permeability [62].

AQPs as potential drug targets

Malfunction of AQPs is associated with several diseases (reviewed in [5, 63–65]). On the one hand, the distorted expression of AQPs was shown to be implicated in diseases such as nephrogenic diabetes insipidus in the kidney [66], inherited cataracts in the eye [67, 68], or deficient secretion of saliva associated with the Sjögren's syndrome [69]. On the other hand, excessive water transport through AQPs was also described as a pathological condition in diseases such as congestive heart failure and cirrhosis [70, 71]; brain [72, 73] and lung [74] edema; glaucoma [63, 64, 75], and lung cancer [76]. In addition to water permeation, AQPs were found to be involved in disorders in glycerol metabolism, such as obesity [77].

Infection diseases, such as malaria, sleeping sickness, Chagas' disease or leishmaniasis, caused by protozoan parasites, are affecting hundreds of millions of individuals in tropical and subtropical regions [23, 24, 78]. The AQPs present in these parasites, in particular the *Plasmodium falciparum* malaria aquaglyceroporin, have been observed to play roles such as water regulation, glycerol uptake or release of toxic products [25, 29, 79–81]. The blockage of such AQPs has been thus suggested as a possibility to interrupt the life cycle of the parasite and prevent its proliferation.

AQPs are therefore promising drug targets for the treatment of pathological disorders associated with water- or glycerol-imbalance, or infectious diseases caused by protozoan parasites.

Molecular dynamics simulations revealed the motion of AQPs

Molecular dynamics simulations bridged the gap between the -static - atomic AQP structures and the macroscopic permeation properties of these proteins, by providing the time evolution of the biomolecular AQP system at atomic resolution [56, 57]. This computational technique gave essential dynamical and energetic information, that is difficult- or even impossible- to obtain with the current available experimental techniques. Quantities such as the single-channel water permeability [82–85] or the free energy for solute permeation [4, 51, 58, 62] could be directly evaluated. With this information at hand, it was possible to shed light onto the molecular mechanisms of permeation [4, 51], selectivity [58, 62] and proton exclusion [59–61] of AQPs (all mechanisms described above). Moreover, it was also possible to investigate the inhibitory effect of a putative AQP1 blocker [86]. Consequently, molecular dynamics has shown to be an extremely useful computational tool, complementary to structural and functional experiments, for the study of AQPs.

Aim of the present thesis

The aim of the present thesis is to expand our understanding on the molecular machinery of aquaporins by using molecular dynamics simulations and related computational techniques. The following specific goals were addressed: first, to provide a quantitative description of the solute permeation through the *Plasmodium falciparum* malaria aquaglyceroporin; second, to investigate the molecular determinants governing aquaporin gating, and third, the study of interactions of aquaporins with other (bio)molecules. The three goals are described in detail in the following sections.

1. Solute permeation through the *Plasmodium falciparum* aquaglyceroporin

The aquaglyceroporin from *Plasmodium falciparum* (PfAQP) is a potential drug target for the treatment of malaria. It efficiently conducts water and other small solutes, and is proposed to participate in several crucial physiological processes during the parasite life cycle. Despite the wealth of experimental data available, a dynamic and energetic description at the single molecule level of the solute permeation through PfAQP has been lacking so far. In chapter 3 this question is addressed by using molecular dynamics simulations. Initially, the water permeation through PfAQP was quantified by computing the single molecule osmotic permeability coefficient. Subsequently, the pore geometry and the energetics for water transport through PfAQP were analyzed by computing the radius and potential of mean force profiles from equilibrium simulations. Finally the energetics of permeation of glycerol and urea was studied by computing potentials of mean force by using the technique of umbrella sampling simulations.

2. The gating mechanism of aquaporins

Regulation of aquaporins has nowadays become an active field of research. Channel gating has emerged as an efficient regulatory mechanism, facilitating a rapid response to external stimuli when other regulatory mechanisms, such as transcriptional regulation or trafficking, are too slow [54]. Accordingly, in chapter 4 molecular dynamics simulations are used to investigate the molecular determinants of gating of the yeast aquaporin (Aqy1), the spinach plant aquaporin (SoPIP1;2), and the human AQP4 (hAQP4):

- Gating of yeast aquaporin mediated by phosphorylation or mechanosensing

The hypothesis that Aqy1 may be gated by either phosphorylation of a serine residue or membrane-mediated mechanical stress is investigated. Phosphorylation may fine tune the water flux during normal conditions of growth, whereas mechanosensitive gating could provide a rapid pressure valve in response to unexpected shocks. In addition, Aqy1 regulation may help yeast to survive rapid freezing and thawing, aiding the organism's quest to adapt and survive.

- Is spinach plant aquaporin gated by phosphorylation or histidine protonation?

For SoPIP2;1, we examine how phosphorylation of two serine residues or protonation of a histidine residue, the two key components of the current proposed gating mechanism for plant aquaporins, affect the water permeation and the structure of the postulated gate of the pore. The possibility that SoPIP2;1 may be gated by membrane-mediated mechanical stress is also analyzed.

- **Voltage-regulation of human AQP4**

A putative voltage-gated mechanism for hAQP4 is explored. Voltage regulation of hAQP4 may have important physiological role in the central nervous system (CNS). There, hAQP4 colocalizes with the Kir4.1 potassium channel, and both proteins have been suggested to work as a water-potassium transport complex [73]. In addition, the possibility of voltage regulation of AQPs has been proposed in a previous simulation study for hAQP1 [87]. Consequently, the possibility that voltage sensitivity may be a general feature of AQPs is analyzed.

3. Interaction of aquaporins with other (bio)molecules

AQPs do not work alone. They interact at different stages with different (bio)molecules such as proteins, lipids and organic compounds. In chapter 5, three important processes where AQPs interact with different types of (bio)molecules are addressed, by means of molecular dynamics simulations. First, the formation and stability of the AQP2-LIP5 complex (aquaporin-protein interactions) is studied. This is a crucial process during the trafficking of AQP2 in renal cells. Second, the localization of DMPC lipids around AQP0 is investigated, being AQP0 an excellent model to study protein-lipid interactions. Third, an approach, combining computational and experimental techniques, to search and refine putative AQP9 blockers is described.

- **The complex formed by AQP2 and the LIP5 protein**

In renal cells, the translocation of AQP2 is a process of vital importance. The lysosomal trafficking regulator protein-5 (LIP5) has been experimentally shown to interact with the C-terminal tail of AQP2, and suggested to facilitate the AQP2 internalization from the plasma membrane [88]. However, the structure of LIP5 coupled to AQP2 (the AQP2-LIP5 complex) has not been determined yet. We use molecular dynamics simulations and related computational techniques to predict putative stable structures of the AQP2-LIP5 complex, mediated by interactions between the C-terminal tail of AQP2 and LIP5, and compatible with the AQP2 tetramer embedded in a lipid bilayer.

- **AQP0: an excellent model to study protein-lipid interactions**

Biological lipid membranes are the natural environment of AQPs and other membrane proteins. Understanding protein-lipid interactions is therefore crucial to elucidate how membrane proteins are stabilized within the lipidic environment and also how they alter the conformational properties of lipids. AQP0 has emerged as an excellent model to study protein-lipid interactions, mainly motivated by recent electron crystallography studies [38, 39] that revealed the structure of some annular lipids around the AQP0 protein. Taking advantage

of this valuable experimental evidence, first, we study how representative the crystallographic lipid structures are of those adopted by annular lipids around aquaporin tetramers immersed at low concentrations in a lipid bilayer. Second, we determine which parts of the lipids (heads or tails) specifically interact with AQP0. Third, we describe how annular lipids behave differently at the AQP0 tetramer surface and within the pure bulk lipid phase.

- Inhibition of AQP9

Specific AQP9 blockers are potential candidates to control disorders in glycerol metabolism [77], and may prevent the propagation of malaria in the initial phases of infection [89, 90]. Here, we present our initial findings in the search and refinement of putative AQP9 blockers, by an interdisciplinary approach combining molecular dynamics simulations, functional assays and molecular docking calculations.

Organization of the thesis

The thesis is organized as follows: Chapter 2 describes the molecular dynamics simulation technique and the methods used to derive observables relevant to AQPs, such as the single-channel water permeability or the free energy for the solute permeation. Chapter 3 presents the study of the *Plasmodium falciparum* aquaglyceroporin (goal 1). Chapter 4 focuses on the gating mechanisms of AQPs (goal 2). Chapter 5 is dedicated to the interactions of AQPs with other (bio)molecules (goal 3). Finally, chapter 6 summarizes the main conclusions and perspectives of the thesis.

Molecular dynamics (MD) simulation is the main computational method used in this thesis. In this chapter, we describe its principles, approximations and algorithms. For a more detailed explanation we refer to several books [91–93], reviews [94–96] and the GROMACS manual [97–99]. In addition, we outline the methods used to derive observables from MD simulations which are relevant to aquaporins.

2.1 Molecular dynamics simulations

MD simulation is a computational method to numerically solve Newton's equations of classical motion of a system of N interacting atoms:

$$m_i \frac{\partial \mathbf{r}_i}{\partial t^2} = \mathbf{F}_i, \quad i = 1 \dots N. \quad (2.1)$$

Here, m_i and \mathbf{r}_i are the mass and position of atom i , respectively, and \mathbf{F}_i is the force acting on atom i . The forces can be expressed as minus the gradient of a potential energy function $V(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N)$, describing all the interatomic interactions:

$$\mathbf{F}_i = -\nabla V(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N). \quad (2.2)$$

Position and velocities are obtained as a function of time for all the N atoms, producing a (MD) trajectory of the system.

In the following subsections we present the approximations in which the MD method is based, the algorithm to solve numerically the equations of motion, and additional algorithms that are implemented to increase the efficiency and accuracy of the method.

2.1.1 Approximations of MD simulations

The MD simulation method is based on the following three approximations: first, the separation of the nuclear and the electronic degrees of freedom (Born-Oppenheimer approximation); second, the classical treatment of the dynamics of atoms, and third, the use of empirical functions (force fields) to describe interatomic interactions.

Born-Oppenheimer

Dynamics of atoms and molecules is governed by quantum mechanics. Within this formalism, the state of the system is represented by a wave function $\psi(\mathbf{R}_e, \mathbf{R}_n)$ that depends on the degrees of freedom of electrons, \mathbf{R}_e , and nuclei, \mathbf{R}_n . This state evolves in time according to the time-dependent Schrödinger equation:

$$i\hbar \frac{\partial}{\partial t} \psi(\mathbf{R}_e, \mathbf{R}_n) = H \psi(\mathbf{R}_e, \mathbf{R}_n), \quad (2.3)$$

where $\hbar = h/2\pi$ is the reduced Planck constant, and H denotes the hamiltonian of the system. The Born-Oppenheimer approximation consists in the separation of the rapid motions of electrons from the slow dynamics of nuclei. The wave function $\psi(\mathbf{R}_e, \mathbf{R}_n)$ is expressed as:

$$\psi(\mathbf{R}_e, \mathbf{R}_n, t) = \psi_n(\mathbf{R}_n, t) \psi_e(\mathbf{R}_e; \mathbf{R}_n), \quad (2.4)$$

where $\psi_n(\mathbf{R}_n)$ and $\psi_e(\mathbf{R}_e; \mathbf{R}_n)$ are the nuclear and electronic wave functions, respectively. Note that the electronic wave function depends on the nuclear degrees of freedom only parametrically. This approximation leads to the separation of equation 2.3 in two equations: a time-dependent Schrödinger equation describing the motion of nuclei and a time-independent eigenvalue Schrödinger equation for the electrons. In consequence, electrons must remain in their ground state and instantaneously adjust their wave function when nuclei positions (taken parametrically) change. In addition, nuclei move in the presence of a stationary potential energy surface created by the electrons, and their positions specify the atom coordinates (\mathbf{r}_i in equation 2.1).

Atom dynamics are described classically

The second approximation is to assume that the dynamic properties of the system are adequately described by laws of classical mechanics. The classical approximation is valid if the

typical separation between atoms is much greater than their mean thermal de Broglie wavelength [100]. For most atoms at room temperature this is usually the case [97]. However, all processes of quantum character such as chemical reactions, tunneling, dynamics of electronically excited states, or interactions of atoms with photons can not be studied with MD simulations.

Force fields: empirical interaction potential functions

The third approximation concerns the potential energy function $V(\mathbf{r}) = V(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N)$, that yields the forces via equation 2.2. $V(\mathbf{r})$ is assumed to be a semi-empirical function, called *force field*, composed of simple terms describing all the essential interatomic interactions, that are easy and computationally cheap to calculate.

The expressions contained in the force field can be categorized in bonded and non-bonded. Bonded terms refer to covalent bonds V_b , angular vibrations V_a , proper dihedrals V_{dih} , and improper dihedrals $V_{imp.dih}$. Non-bonded terms correspond to electrostatic Coulomb interactions V_{coul} , and short-range dispersion interactions V_{LJ} . A typical force field function is given by [97]

$$\begin{aligned}
 V(\mathbf{r}) &= V_b + V_a + V_{dih} + V_{imp.dih} + V_{coul} + V_{LJ} & (2.5) \\
 &= \sum_{\text{bonds } i} \frac{k_i}{2} (l_i - l_{i,0})^2 \\
 &= + \sum_{\text{angles } i} \frac{f_i}{2} (\varphi_i - \varphi_{i,0})^2 \\
 &= + \sum_{\text{dihedrals } i} \frac{V_i}{2} [1 + \cos(n\phi_i - \phi_{i,0})] \\
 &= + \sum_{\text{imp.dih } i} \kappa_i (\xi_i - \xi_{i,0})^2 \\
 &= + \sum_{\text{pairs } i,j} 4\epsilon_{i,j} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{4\pi\epsilon_0\epsilon_r r_{ij}} .
 \end{aligned}$$

Here, bonds, angles, dihedrals and improper dihedrals are modeled as harmonic potentials. Improper dihedrals are included to keep some groups planar (e.g. aromatic rings). Proper dihedrals are used to allow different torsional orientations of groups such as alkanes. Non-bonded interactions are evaluated between pairs of atoms. Short-range repulsive and attractive dispersion interactions are modeled by a Lennard-Jones (LJ) potential with the parameters σ_{ij} and ϵ_{ij} determining the strength of the potential. Finally, electrostatic coulomb interactions are considered between partial charges q_i and q_j of atoms i and j , respectively.

Parameters of a force field are usually calculated either from *ab initio* quantum mechanical calculations or by fittings to reproduce experimental thermodynamics properties such as solvation free energies of aminoacids. Several force fields are available in the literature, e.g. OPLS-AA [101, 102], amber [103], GROMOS [104] and CHARMM [105, 106]. For the present thesis the OPLS-AA and the amber99SB force fields were used.

2.1.2 Numerical solution of the equations of motion

Several algorithms have been proposed to numerically integrate equation 2.1, and thereby obtain positions and velocities of all the atoms of the system as a function of time. In this thesis, the *leap frog* algorithm [107] was used. The algorithm scheme looks as follows: First, it is assumed that velocities \mathbf{v}_i and positions \mathbf{r}_i are known at times $t - \frac{\Delta t}{2}$ and t , respectively. Second, forces are computed at time t using equations 2.2 and 2.5. Third, velocities and positions are updated according to the leap frog iterative formulas [107]

$$\mathbf{v}_i \left(t + \frac{\Delta t}{2} \right) = \mathbf{v}_i \left(t - \frac{\Delta t}{2} \right) + \frac{\mathbf{F}_i(t)}{m_i} \Delta t, \quad (2.6)$$

$$\mathbf{r}_i(t + \Delta t) = \mathbf{r}_i(t) + \mathbf{v}_i \left(t + \frac{\Delta t}{2} \right) \Delta t. \quad (2.7)$$

The integration continues by repeating these three steps iteratively. The leap frog algorithm is of third order in positions and velocities, and is equivalent to the Verlet [108] algorithm.

2.1.3 Simulation details

In practice, additional algorithms are implemented together with the leap frog algorithm to increase the efficiency and accuracy of the method. In this section we briefly outline these additional algorithms.

The time step Δt is usually chosen such that at least 5 integration steps are performed within the smallest oscillation period considered in the simulation. Bond-stretching and bond-angle vibrations are typically the degrees of freedom with smallest oscillation periods, allowing a time step of 1 fs. In the simulations presented in this thesis, bond lengths are constrained allowing a time step of 2 fs. The Settle algorithm [109] was used to constrain bond lengths and angles of water molecules and Lincs [110] was used to constrain all other bond lengths. The next fastest degrees of freedom are bond angles involving hydrogen atoms which were removed by using the virtual interaction-sites algorithm [111], allowing a 4 fs time step.

Periodic boundary conditions were implemented to minimize finite size effects. The simulation box is repeated periodically in all three dimensions. Consequently, boundaries are removed and atoms exiting one simulation box enter the neighbour one. Recall that the errors introduced

by finite size effects are now replaced by the errors of a periodic system. However, possible artifacts due to the periodicity are expected to be smaller than the errors induced by an unnatural vacuum boundary [97].

Evaluation of the non-bonded interaction term in the force field (equation 2.5) is the most expensive part of the integration step, because it contains a double sum over all the atoms of the system. To improve this part of the algorithm, short-range Lennard-Jones interactions were only considered within a certain cutoff distance, typically 1.0 nm. In addition, a search of nearest neighbours was implemented, and it was not updated at every simulation step. The neighbour lists were updated every 20 fs (corresponding to 5 and 10 integration steps of 4 fs and 2 fs, respectively).

A cutoff scheme can not be applied for the Long-range electrostatic interactions because it induces severe artifacts in the dynamics of the system [112, 113]. To solve this issue, the particle-mesh Ewald (PME) method [114, 115] was implemented. The method is based on the Ewald sums originally formulated to compute long-range interaction in periodic systems. The electrostatic Coulomb term is separated in short-range and long-range contributions, calculated in the direct and the reciprocal space, respectively. The long-range contribution is recovered back to the real direct space by applying a fast Fourier transformation. Direct calculation of the electrostatic term scales with N^2 , N being the number of charges, whereas the PME method improves the scaling to $N \log N$.

Finally, the system was assumed to be in contact with a reservoir in the NPT (isobaric-isothermal) canonical ensemble. Temperature and pressure are therefore kept constant by coupling the system to a thermostat and barostat, respectively. The algorithms for thermostats [116–121] employ different methodologies such as the rescaling of the velocities in each step to gradually converge to a reference temperature or the inclusion of stochastic terms (noise) in the equations of motion. Accordingly, in the algorithms for barostats [116, 122] the positions of atoms are rescaled to approach towards the reference pressure or additional terms are added to the equations of motion. In this thesis, the temperature was kept constant by coupling the system to either a Berendsen [116] or a velocity rescaling thermostat [117] at 300 K. The pressure was kept constant by coupling the system to either a Berendsen [116] or a Parrinello-Rahman barostat [122] at 1 bar.

In the present thesis, the MD simulations were carried out using the GROMACS simulation software [98, 99]. All the mentioned methods and algorithms are already implemented within this package, allowing for the efficient simulation of aquaporins in their native environment, embedded in a membrane surrounded by explicit water molecules. For an estimation of the performance of the method, the simulation of 100 ns (25×10^6 integration steps of 4 fs)

of an aquaporin system containing approximately 10^5 atoms lasted approximately 10 days, using 16 cpu cores working in parallel. The PyMOL [123] and VMD [124] packages were used for the visualization of the simulation system and the MD trajectories.

2.2 Observables derived from MD simulations

MD simulations and related computational techniques have enormously contributed to the understanding of the mechanism of permeation, specificity and regulation of aquaporins, by providing dynamical and energetic information at the atomic level, that is usually difficult to access experimentally. In the following subsections we describe the methods employed to derive observables from MD simulations which are relevant for aquaporins.

2.2.1 Water permeability coefficients

Let us to consider two water compartments conected by a channel, and certain solute deposited in each compartment. A solute concentration difference, ΔC_s , between these two comparments creates a hydrostatic pressure difference. A net water flux j through the channel is induced in response to the hydrostatic pressure difference. The channel osmotic permeability, p_f , is defined as minus the ratio between j and ΔC_s [3],

$$j = -p_f \Delta C_s. \quad (2.8)$$

The p_f is therefore and intrinsic property of the channel that quantifies how fast water permeation occurs in response to external solute concentration gradients.

Several methods have been proposed to derive the p_f from either equilibrium or non-equilibrium molecular dynamics simulations. In non-equilibrium methods [83], hydrostatic pressure differences are directly applied between the two water compartments. Alternatively, equilibrium methods efficiently derive the p_f from the random correlated motions of waters inside the channel in the absence of external driving forces. Equilibrium methods are based on rate theory [57, 82], collective diffusion models [84] or matrix analysis [85].

In this thesis, the p_f was independently calculated for every aquaporin monomer (that is a water channel) by using the equilibrium collective diffusion model proposed by Zhu *et al.* [84], explained briefly as follows: A collective variable n is defined as

$$dn = \sum_i dz_i / L, \quad (2.9)$$

where dz_i are the displacements along the pore coordinate of the water molecules inside the channel (labeled as i) during a time dt , and L is the length of the channel. The variable n

obeys an Einstein relation,

$$\langle n^2(t) \rangle = 2D_n t, \quad (2.10)$$

and the proportionality constant, $2D_n$, is related with the permeability coefficient [84]:

$$p_f = v_w D_n, \quad (2.11)$$

where $v_w \simeq 30 \text{ \AA}^3$ is the volume occupied by one water molecule. Thus, the p_f can be assessed by monitoring the displacements dz_i of the water molecules inside the channel during the MD simulations.

2.2.2 Potentials of mean force

An important energetic quantity that can be derived from MD simulations is the so-called potential of mean force (PMF). The PMF was originally introduced by Kirkwood [125] to study chemical potentials of fluid mixtures. The following paragraphs briefly describe the physical basis of the PMF, based on the thesis by Hub, H. [87] and Portella, G. [126]. For a N particle system with degrees of freedom $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{3N}$, with \mathbf{x}_1 being certain reaction coordinate (e.g the pore axis of a channel), the PMF $W(\mathbf{x}_1)$ is defined as [127]

$$e^{-\beta W(\mathbf{x}_1)} = C \int e^{-\beta V(\mathbf{x}_1, \dots, \mathbf{x}_{3N})} d\mathbf{x}_2 \dots d\mathbf{x}_{3N}. \quad (2.12)$$

Here, V is the potential energy; C is a constant, and $\beta = 1/k_B T$, where k_B is the Boltzmann constant and T the temperature. Note that the integration has been carried out over all the degrees of freedom except \mathbf{x}_1 . The name "*potential of mean force*" is due to the fact that the derivative of $-W(\mathbf{x}_1)$ with respect to \mathbf{x}_1 , $-\frac{\partial W(\mathbf{x}_1)}{\partial \mathbf{x}_1}$, is the average force that a particle "feels" at a fixed position along the reaction coordinate \mathbf{x}_1 :

$$\begin{aligned} -\frac{\partial W(\mathbf{x}_1)}{\partial \mathbf{x}_1} &= \frac{\int \left(-\frac{\partial V(\mathbf{x}_1, \dots, \mathbf{x}_{3N})}{\partial \mathbf{x}_1} \right) e^{-\beta V(\mathbf{x}_1, \dots, \mathbf{x}_{3N})} d\mathbf{x}_2 \dots d\mathbf{x}_{3N}}{\int e^{-\beta V(\mathbf{x}_1, \dots, \mathbf{x}_{3N})} d\mathbf{x}_2 \dots d\mathbf{x}_{3N}} \\ &= \left\langle -\frac{\partial V(\mathbf{x}_1, \dots, \mathbf{x}_{3N})}{\partial \mathbf{x}_1} \right\rangle \\ &= \langle F(\mathbf{x}_1) \rangle. \end{aligned} \quad (2.13)$$

The PMF is a constrained free energy

The PMF can also be considered as a constrained free energy as shown in the following paragraphs. The free energy \mathcal{W} is related with the partition function Z according to [100]

$$\beta \mathcal{W} = -\ln Z. \quad (2.14)$$

If the system is kept at constant volume (NVT ensemble), \mathcal{W} is called the Helmholtz free energy F , whereas, if it is kept at constant pressure (NPT ensemble), \mathcal{W} is referred to the Gibbs free energy G . The partition function Z is defined as

$$Z = \int e^{-\beta H} d^{3N} \mathbf{x} d^{3N} \mathbf{p}, \quad (2.15)$$

where $H = T + V$ denotes the hamiltonian of the system, that is the sum of the kinetic T plus the potential energy V . The integration is performed over the whole phase space, covering all possible microstates accessible to the system.

Now, let us consider a certain observable $\mathcal{A}(\mathbf{x}, \mathbf{p})$ constrained to a fixed value ξ . The partition function constrained to ξ is thus given by

$$Z(\xi) = \int \delta(\mathcal{A}(\mathbf{x}, \mathbf{p}) - \xi) e^{-\beta H} d^{3N} \mathbf{x} d^{3N} \mathbf{p}, \quad (2.16)$$

with δ denoting the Dirac delta function. Accordingly, the constrained free energy reads

$$\beta \mathcal{W}(\xi) = -\ln Z(\xi) \quad (2.17)$$

$$= -\ln \int \delta(\mathcal{A}(\mathbf{x}, \mathbf{p}) - \xi) e^{-\beta H} d^{3N} \mathbf{x} d^{3N} \mathbf{p}. \quad (2.18)$$

In the particular case that \mathcal{A} represents the reaction coordinate \mathbf{x}_1 , the constrained free energy $\mathcal{W}(\xi)$ yields

$$e^{-\beta \mathcal{W}(\xi)} = C \int e^{-\beta V(\xi, \mathbf{x}_2, \dots, \mathbf{x}_{3N})} d\mathbf{x}_2 \dots d\mathbf{x}_{3N}. \quad (2.19)$$

Here, the constant C is the term associated to the kinetic energy T :

$$C = \int e^{-\beta T(\mathbf{p}_1, \dots, \mathbf{p}_{3N})} d^{3N} \mathbf{p}. \quad (2.20)$$

Equation 2.19 corresponds the definition of the PMF given in equation 2.12, showing that the PMF can also be considered as a constrained free energy.

The constrained partition function also provides the probability for the system to be in a state with the observable $\mathcal{A} = \xi$:

$$P(\xi) = \frac{Z(\xi)}{Z}, \quad (2.21)$$

$$= \frac{1}{Z} \int \delta(\mathcal{A}(\mathbf{x}, \mathbf{p}) - \xi) e^{-\beta H} d^{3N} \mathbf{x} d^{3N} \mathbf{p}. \quad (2.22)$$

The free energy can be obtained in terms of the probability $P(\xi)$ by replacing equation 2.21 in equation 2.17:

$$\mathcal{W}(\xi) = \mathcal{W}(\xi^*) - K_B T \ln \frac{P(\xi)}{P(\xi^*)}. \quad (2.23)$$

Here, the free energy $\mathcal{W}(\xi)$ is relative to $\mathcal{W}(\xi^*)$, that is a value associated to a particular (arbitrary) chosen reference value ξ^* . $\mathcal{W}(\xi)$ can be thus determined from MD simulations by computing the probability $P(\xi)$.

Computing the PMF for water permeation through water channels from equilibrium MD simulations

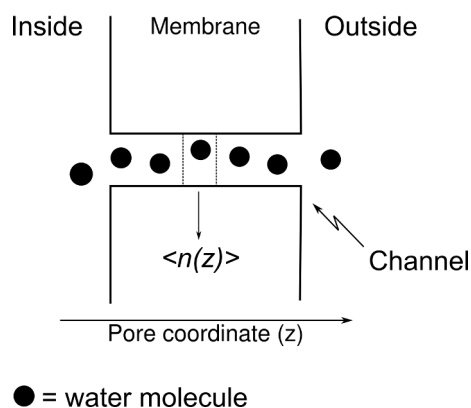


Figure 2.1. Schematic representation of a water channel separating two water compartments. The PMF for water permeation through the channel is derived from MD simulations by using equation 2.23. The axis along the pore (denoted as z) was assumed to be the reaction coordinate ξ , and the probability $P(\xi = z)$ was estimated as the average (over the whole trajectory) of the number of water molecules at different positions along the channel, $\langle n_i(z) \rangle$.

The potential of mean force (PMF) for water permeation through aquaporins was calculated from equilibrium MD simulations as explained as follows. The axis along the pore (denoted as z) was assumed to be the reaction coordinate ξ , and $P(\xi = z)$ was estimated as the average (over the whole trajectory) of the number of water molecules at different positions along the channel, $\langle n_i(z) \rangle$ (see figure 2.1). Simulations were carried out in the isobaric ensemble, thus the PMF refers to the Gibbs free energy. Aquaporins form tetramers, where each monomer constitutes a water channel, thus, for the i -th pore, equation 2.23 reduces to

$$G_i(z) = -k_B T \ln \langle n_i(z) \rangle + G^*, \quad (2.24)$$

where G^* is a calibration constant chosen such that the free energy at the water compartments is zero. The number of water molecules is referred to a cylinder that is aligned with the pore coordinate. This reference state is chosen to allow a direct comparison of the free energy of water permeation through the channel and the alternative pathway through the membrane. As a consequence, the entropy at the bulk regions is underestimated, and consequently, the free energy at the channel relative to the bulk is also underestimated. To relate $G_i(z)$ with the area of one aquaporin monomer and account for such underestimations, a trapezoidal correction was applied in the entrance and exit regions. In consequence, the final PMF refers to a density of one channel per membrane area occupied by an aquaporin monomer [58]. The correction reads

$$\Delta G_{\text{cyl}} = k_B T \ln(A_{\text{mono}}/A_c) \quad , \quad (2.25)$$

where A_{mono} denotes the membrane cross section area of the aquaporin monomer and A_c the cross section area of the cylinder ($\pi \times 0.25 \text{ nm}^2$). Prior to the analysis the protein monomers were superimposed to a reference structure. The effective PMF, $G_{\text{eff}}(Z)$, was computed by

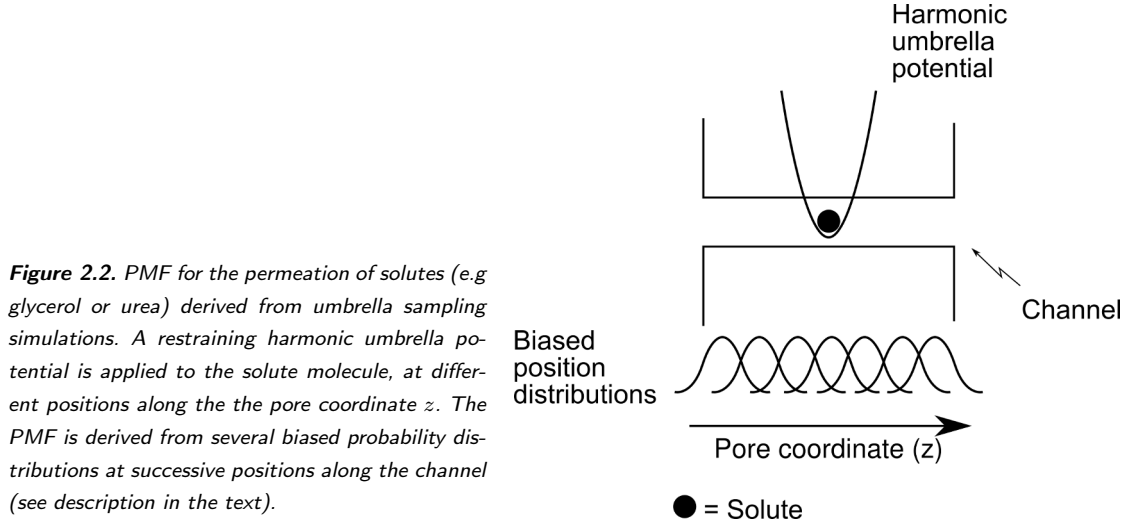
combining the four monomer $G_i(z)$ as

$$\exp(-G_{eff}(Z)/k_B T) = \frac{1}{4} \sum_{i=1}^4 \exp(-G_i(Z)/k_B T) = \frac{1}{4} \sum_{i=1}^4 \langle n_i(z) \rangle, \quad (2.26)$$

and the error of G_{eff} was estimated by propagating the standard errors of $\langle n_i(z) \rangle$.

PMF for the permeation of other solutes derived from umbrella sampling simulations

For other solutes such as glycerol or urea, it is not practical to calculate the PMF from equilibrium MD simulations [127]. Unlike water, these solutes may not sample - simply by diffusion - all possible positions along the pore coordinate in the available computer time (~ 100 ns). This is attributed, first, to high energy barriers that solutes may encounter inside the channel, and second, to the fact that typical physiological solute concentrations (~ 500 mM) are very low compared to the water concentration (55.5 M).



Alternatively, the PMF for the permeation of such solutes was calculated by employing the so-called umbrella sampling method. Here, all positions along the reaction coordinate are efficiently sampled by applying an additional restraining harmonic umbrella potential to the solute molecule, at different successive positions ξ_i along the reaction coordinate ξ (see figure 2.2). The harmonic potential for the i -th position reads

$$\omega_i(\xi) = \frac{1}{2} k (\xi - \xi_i)^2, \quad (2.27)$$

where k is the force constant. The biased probability for the system for the system to be in a state with the observable $\mathcal{A} = \xi$ is defined according to equation 2.21 by

$$P_i^{\text{bias}}(\xi) = \frac{\int \delta(\mathcal{A}(\mathbf{x}) - \xi) e^{-\beta(H+\omega_i)} d^{3N} \mathbf{x} d^{3N} \mathbf{p}}{\int e^{-\beta(H+\omega_i)} d^{3N} \mathbf{x} d^{3N} \mathbf{p}}. \quad (2.28)$$

Here, we have assumed that $A(x)$ only depends on the coordinates. Let us assume that the observable A is the pore coordinate along the channel z . Making use of the Dirac delta function properties modifies equation 2.28 to

$$\begin{aligned} P_i^{\text{bias}}(\xi) &= e^{-\beta\omega_i(\xi)} \frac{\frac{1}{Z} \int \delta(\mathcal{A}(\mathbf{x}) - \xi) e^{-\beta H} d^{3N} \mathbf{x} d^{3N} \mathbf{p}}{\frac{1}{Z} \int e^{-\beta(H+\omega_i)} d^{3N} \mathbf{x} d^{3N} \mathbf{p}}, \\ &= e^{-\beta\omega_i(\xi)} \frac{P(\xi)}{\langle e^{-\beta\omega_i(\xi)} \rangle}. \end{aligned} \quad (2.29)$$

Here we made use of the fact that the numerator corresponds to the unbiased probability $P(\xi)$ (see equation 2.21) and the denominator is the ensemble average of $e^{-\beta\omega_i(\xi)}$. Plugging $P(\xi)$ from equation 2.29 into equation 2.23 yields the unbiased PMF for the i -th umbrella window

$$\mathcal{W}_i(\xi) = W(\xi^*) - K_B T \ln \frac{P_i^{\text{bias}}(\xi)}{P(\xi^*)} - \omega_i(\xi) + F_i, \quad (2.30)$$

with the constant F_i defined by

$$F_i = -K_b T \ln \langle e^{-\beta\omega_i(\xi)} \rangle. \quad (2.31)$$

The PMF of the i -th window is thus obtained by computing the biased probability $P_i^{\text{bias}}(\xi)$ and the constant F_i . In order to reconstruct the overall PMF from the different umbrella windows, the weighted histogram analysis method (WHAM) [128] was used. WHAM is nowadays a well established efficient method to derive the PMF from biased (umbrella sampling) MD simulations [127]. The idea of WHAM is to minimize the error of $P(\xi)$ by combining the probability distributions of N_w overlapping umbrella windows. A more detailed explanation can be found in the original publication [128] and a book on the subject [129]. According to equation 2.29 the unbiased probability distribution for the i -th window is given by

$$P_i^{\text{unbiased}} = e^{-\beta(F_i - \omega_i(\xi))} P_i^{\text{bias}}(\xi). \quad (2.32)$$

The overall probability distribution is expressed as a weighted sum of N_w unbiased probability distributions:

$$p(\xi) = \sum_{i=1}^{N_w} w_i P_i^{\text{unbiased}}(\xi), \quad (2.33)$$

with w_i denoting the weights. The error of such a probability is minimized with respect to the weights, yielding the WHAM equations

$$p(\xi) = \frac{\sum_{i=1}^{N_w} n_i P_i^{\text{bias}}(\xi)}{\sum_{i=1}^{N_w} n_i e^{-\beta\omega_i(\xi) - F_i}}, \quad (2.34)$$

$$e^{-\beta F_i} = \int e^{-\beta\omega_i(\xi)} p(\xi) d\xi. \quad (2.35)$$

Here, n_i is the number of independent data points used to construct $P_i^{bias}(\xi)$. WHAM equations are solved self-consistently, by performing several iterations steps until $p(\xi)$ reaches a converged value.

2.2.3 Principal component analysis

Principal component analysis (PCA) has shown to be a useful method to detect relevant degrees of freedom of biomolecules, from their trajectories obtained e.g with MD simulations. The method separates the configurational space into few collective essential degrees of freedom, covering most of the positional fluctuations, and the remaining degrees of freedom, which represent less-relevant fluctuations that can be discarded for many purposes [130]. These essential degrees of freedom, also called principal components, correspond to collective motions of one or various parts of the biomolecule, and often turned out to be relevant motions for the function of the biomolecule [130]. The method consists of the calculation and diagonalization of the covariance matrix of the positions of a set of atoms inside the biomolecule. The following explanation is based on the PCA original paper by Amadei, A. *et al.* [130] and the thesis by Kubitzki, M. [131].

Let us consider certain part of a biomolecule, composed of N atoms, as the region of interest for the PCA analysis. Let \mathbf{x} represent the vector with the coordinates of these atoms, $\mathbf{x} = (x_1, \dots, x_{3N})$, and $x(t)$ denote their trajectory, derived from e.g. an MD simulation. The covariance matrix of the positions is defined by

$$C = \langle (\mathbf{x} - \langle \mathbf{x} \rangle)(\mathbf{x} - \langle \mathbf{x} \rangle)^T \rangle, \quad (2.36)$$

where $\langle \rangle$ denotes ensemble average. The symmetric matrix C can always be diagonalized by applying an orthogonal coordinate transformation T into a diagonal matrix Λ , with the diagonal elements being the eigenvalues λ_i of C ,

$$\Lambda = T^T C T \quad \text{or} \quad C = T \Lambda T^T. \quad (2.37)$$

The i th column of $T = (\mu_1, \dots, \mu_{3N})$ contains the normalized eigenvector (principal component) μ_i of C corresponding to λ_i . The eigenvalues correspond to the mean square positional fluctuation along the respective eigenvector, and therefore contain the contribution of each principal component to the total fluctuation. Most of the fluctuations are usually concentrated in few eigenvectors corresponding to the largest eigenvalues, and those are considered as the essential degrees of freedom. The trajectory $\mathbf{x}(t)$ can be projected onto a particular eigenvector μ_i :

$$p_i(t) = \mu_i \cdot (\mathbf{x}(t) - \langle \mathbf{x} \rangle) \quad , p_i(t) \in \mathbb{R}. \quad (2.38)$$

Usually, the conformational ensemble sampled in the trajectory is projected onto two or three principal eigenvectors, to compare with other ensembles in this reduced essential subspace. In addition, projections back into Cartesian space can be used to visualize the atomic displacements associated with a particular eigenvector,

$$\mathbf{x}_i'(t) = p_i(t)\mu_i + \langle \mathbf{x} \rangle \quad (2.39)$$

Dynamics and energetics of solute permeation through the
Plasmodium falciparum aquaglyceroporin

3.1 Introduction

The parasite *Plasmodium falciparum* is responsible for the most lethal form of the malaria disease [78]. It expresses one aquaglyceroporin (PfAQP) [25], that has been proposed to play crucial physiological roles during the parasitic life cycle [25, 79], and is therefore a potential antimalarial drug target [23, 24, 132].

PfAQP has been intensively studied experimentally over the last decade. Functional studies[25] established that PfAQP conducts water and glycerol at high permeability rates, comparable to the rates of the human aquaporin 1 (hAQP1) in the case of water. These experiments, therefore, suggested that PfAQP is implicated in tasks such as osmotic stress regulation, and glycerol uptake for lipid synthesis and oxidative stress regulation[25]. In addition, PfAQP was found to be permeable to urea [25], ammonia [80], other polyols up to five carbons long [25], carbonyl compounds [79] and arsenite [25], suggesting that PfAQP may also be involved in the release of toxic products. The biological role of aquaglyceroporins for the *Plasmodium* parasites was further investigated in experiments deleting the orthologue of PfAQP in the rodent malaria parasite, *Plasmodium berghei* (PbAQP), that showed a slower proliferation of the parasite in infected mice [133]. Furthermore, mutation studies demonstrated that a glutamate residue located at the C loop (Glu125) near the conserved Arg196 is critical for water permeation [134].

The structure of PfAQP was recently determined by X-ray crystallography at 2.05 Å resolution [55]. It revealed the 3D architecture of the channel and provided molecular insights into the efficient water and glycerol permeation. PfAQP has the same fold as compared to other members of the family of aquaglyceroporins[8]. The pore geometry is remarkably similar to the *Escherichia coli* glycerol facilitator (GlpF)[55], the closest homologous aquaglyceroporin (with a known structure) with a sequence similarity of 50% [25]. PfAQP arranges in a tetrameric structure, where each monomeric unit constitutes a conduction pore. The conduction pore contains two constriction regions: one located near the NPA motifs, that are replaced in PfAQP by unusual NLA and NPS motifs, and a second one at the ar/R region, where Arg196 is located facing the hydrophobic Trp50 and Phe190 residues [55]. In addition, the C loop is anchored to the extracellular vestibule with Glu125 located in proximity to Arg196, highlighting on the importance of this region for the solute permeation [55], as previously hypothesized from functional studies [134].

Despite the wealth of data available, a dynamical and energetic description at the single-molecule level of solute permeation through PfAQP is lacking so far. Furthermore, a systematic comparison of PfAQP with hAQP1 and GlpF (the most studied aquaporin and aquaglyceroporin, respectively), in terms of the solute permeation and their dynamical causes, is incomplete. In this chapter we address these two questions employing molecular dynamics simulations, based on the X-ray crystallographic structure of PfAQP. Initially, the water permeation through PfAQP was quantified by computing the single molecule osmotic permeability coefficient. Subsequently, the pore geometry and the energetics for water transport through PfAQP were analyzed by computing the radius and potential of mean force profiles from equilibrium simulations. Finally the energetics of permeation of glycerol and urea was studied by computing potentials of mean force by using the technique of umbrella sampling simulations.

3.2 Theory and methods

3.2.1 Equilibrium molecular dynamics simulations

Equilibrium molecular dynamics simulations were carried out starting with the aquaporin tetramer in a fully solvated Dipalmitoylphosphatidylcholine (DPPC) lipid bilayer (Fig. 3.1). Three independent simulations of pfAQP aquaporin (labeled from I to III) were carried out: a control simulation of the wild-type form (I); a mutant where Glu125 (located at the C loop) was mutated into serine (II), and a second mutant where Arg196 (located at the pore face) was replaced by alanine (III). The latter two simulations were performed to study the effect of

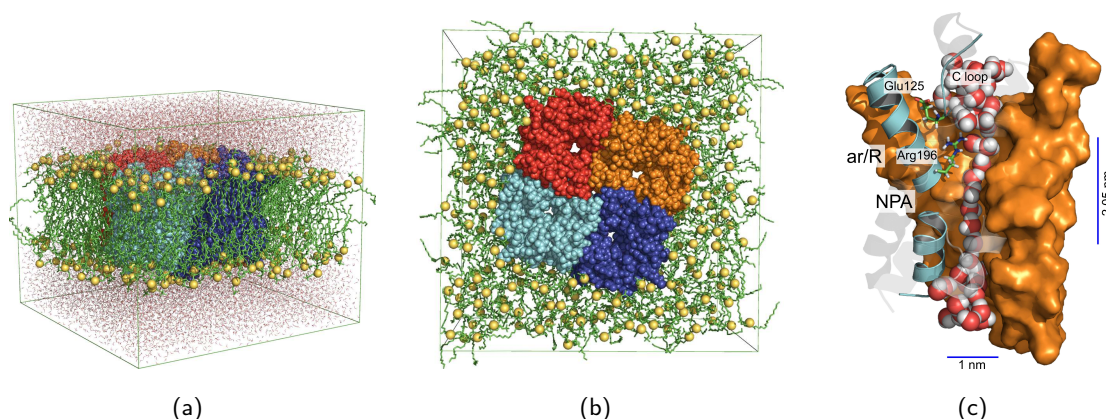


Figure 3.1. Molecular dynamics simulations of PfaAQP. Side (a) and top views (b) of the simulation box illustrating the tetramer (red, orange, cyan and blue), fully embedded in a DPPC lipid bilayer (yellow phosphorous atoms and green tails) and solvated by water (red, white). (c) Conduction pore showing the helices HB and HE; the C loop (cyan), and Glu125 and Arg196 (green). The blue lines indicate the region considered to compute the water permeability coefficient p_f .

these two residues on the water permeation (Fig. 3.1(c)). In addition, simulations of hAQP1 (IV) and GlpF (V) aquaporins were also carried out for comparison.

The simulation boxes contain the protein tetramer, 265 DPPC lipids (263 in simulation V) and around 21500 SPC/E water molecules [135]. The PfaAQP and GlpF structures were taken from the Protein Data Bank (PDB ID codes 3C02 [55] and 1FX8 [50], respectively). The starting structure of hAQP1 was modeled based on the x-ray structure of bovine AQP1 (PDB ID code 1J4N) [42] by mutating differing residues by using the WHAT IF modeling software [136]. The tetramer was inserted into the lipid bilayer by using the `g_membed` software [137]. Crystallographic water molecules were kept in the structures and ions were added to neutralize the simulation systems. The amber99SB all-atom force field [103] was used for the protein, and lipid parameters were taken from Berger *et al.* [138]. The simulations were carried out using the GROMACS simulation software [98, 99]. Long-range electrostatic interactions were calculated with the particle-mesh Ewald method [114, 115]. Short-range repulsive and attractive interactions were described by a Lennard-Jones potential, which was cut off at 1.0 nm. The Settle algorithm [109] was used to constrain bond lengths and angles of water molecules and Lincs [110] was used to constrain all other bond lengths. The fastest angular degrees of freedom involving hydrogen atoms were removed by using the virtual interaction-sites algorithm [111], allowing a time step of 4 fs. The temperature was kept constant by coupling the system to a velocity rescaling thermostat [116, 117] at 300 K with a coupling constant $\tau = 0.5$ ps. The pressure was kept constant by coupling the system to a semiisotropic Parrinello-Rahman barostat [122] at 1 bar with a coupling constant of $\tau = 5.0$ ps.

All simulations were equilibrated for 4 ns before production. During this equilibration time the coordinates of the protein were harmonically restrained, with a harmonic force constant of $1000 \text{ kJmol}^{-1}/\text{nm}^2$. The simulation length of the production runs was 100 ns discarding the first 10 ns as equilibration.

3.2.2 Water permeability calculations

The single-channel osmotic permeability, p_f , was chosen to quantify the water permeation through the simulated aquaglyceroporins. The p_f was independently calculated for every monomer, based on the collective diffusion model [84] described in section 2.2.1. Water molecule displacements, dz_i , were computed every 2 ps, within a cylindrical region (of length $L = 2.05 \text{ nm}$ and radius $r = 0.5 \text{ nm}$) centered at the pore axis and spanning -0.7 nm down and 1.35 nm up from the NPA region (Fig. 3.1(c)). $\langle n^2(t) \rangle$ was obtained by averaging over 450 time windows of 200 ps length each. Finally the p_f was obtained from the slope of the curve $\langle n^2(t) \rangle$ versus time. An effective p_f value was obtained by averaging the values of the four monomers and the error was estimated as the standard error.

3.2.3 Pore dimensions

The geometry of the channels was monitored by computing pore radius profiles with the HOLE software [139], averaging over snapshots taken every 50 ps, yielding a standard error always smaller than 0.4 \AA

3.2.4 Potential of mean force for water

The potential of mean force (PMF) for water was calculated to identify rate limiting regions and relevant water-protein and water-water interaction sites inside the pore. It was independently calculated for every monomer by computing the the average (over the whole trajectory) of the number of water molecules at the z position along the pore (see section 2.2.2). The number of water molecules is counted within a cylinder (of radius 0.5 nm) that is aligned with the pore coordinate (Fig. 3.1(c)). Therefore, the entropy at the bulk regions is underestimated, and consequently, the free energy at the channel relative to the bulk is also underestimated. A trapezoidal correction was applied in the entrance and exit regions to relate $G(z)$ with the area of one aquaporin monomer and account for this underestimation (see equation 2.25). In consequence, the final PMF refers to a density of one channel per membrane area occupied by an aquaporin monomer[58]. The computed correction was 6.1 kJ/mol for pfAQP, 6.0 kJ/mol for hAQP1 and 6.2 kJ/mol for GlpF. Prior to the analysis the protein

monomers were superimposed to a reference structure. The effective PMF, $G_{eff}(Z)$, was computed by combining the four monomer $G_i(z)$ using equation 2.26 and the error of G_{eff} was estimated by propagating the standard errors of $\langle n_i(z) \rangle$, yielding a maximum uncertainty of approximately 1.5 kJ/mol.

3.2.5 Umbrella sampling simulations

The starting frames for the umbrella simulations were taken from a 10 ns equilibrium simulations of PfAQP. The simulation system contained the PfAQP tetramer, 264 POPE lipids, 20484 TIP4P water molecules [140], and four chloride ions. Protein and glycerol interactions were described using the OPLS all-atom force field [101, 102], lipid parameters were taken from Berger *et al.* [138], and urea parameters from refs. [141, 142]. All simulation parameters were chosen as described in the section 3.2.1, except that hydrogen atoms were not described as virtual sites, requiring a time step of 2 fs. The PfAQP channel was divided into 0.25 Å (1 Å for urea) wide equidistant sections parallel to the membrane with the center of each section representing an umbrella center. Subsequently, the solute molecules (glycerol or urea) were placed into the channel at the umbrella center. To enhance sampling, several solute molecules were placed in each pore keeping a distance of 20 Å between the solute molecules. Water molecules that overlapped with the solute were removed. The central carbon atom of glycerol or urea was restrained in the z direction by a harmonic umbrella potential ($k = 1000 \text{ kJmol}^{-1}/\text{nm}^2$). In addition, the solute molecules were restrained into cylinders that were centered along the respective channel. Accordingly, a flat-bottom quadratic potential in the xy -plane was applied on the restrained atom (radius $r_c = 6 \text{ Å}$, $k_c = 400 \text{ kJmol}^{-1}/\text{nm}^2$). After energy minimization, each system was simulated for 2 ns (4 ns for urea PMFs).

After removing the first 500 ps (1 ns for urea) for equilibration from each umbrella simulation, 1440 umbrella histograms (360 for urea) were extracted from the z -coordinate of the restrained atom. Subsequently, the umbrella positions were corrected with respect to the center of mass of the corresponding monomer. This procedure avoids a possible unphysical flattening of the PMF due to fluctuations of the PfAQP monomers within the tetramer. Visual inspection of the umbrella histograms showed that multiple histograms overlapped at each z coordinate. The PMF was calculated using a cyclic implementation of the weighted histogram analysis method (WHAM) [128]. The WHAM procedure incorporated the integrated autocorrelation times (IACT) of the umbrella windows. The IACTs were derived by fitting a double exponential to the autocorrelation function of each window, allowing one to analytically compute the IACTs. Because the IACT is subject to large uncertainty in case of limited sampling (such as inside the channel), we have subsequently smoothed the IACT along the reaction coordinate using

a moving average filter with a width of 5 Å.

Due to the cylindric flat-bottom restraint the umbrella samplings yield a PMF which refers to a channel density of one channel per cross section area of the cylinder. We also corrected the PMF by a trapezoidal correction like the water PMF above. In this case, A_c was chosen such that the entropy of the solute in the flat-bottom quadratic potential equals the entropy of a solute in a cylindrical well-potential of area A_c . We found that this condition approximately holds if A_c is computed via $A_c = \pi(r_c + 2\sigma_c)^2$ where $\sigma_c = (k_B T/k_c)^{1/2}$ is the width of the Gaussian-shaped solute distribution at the edge of the flat-bottom quadratic potential. A_{mono} was estimated to equal 10.7 nm² and A_c was computed to 1.80 nm², yielding a correction of $\Delta G_{\text{cyl}} = 4.4$ kJ/mol. The statistical uncertainty of the PMFs was estimated using bootstrap analysis as described before [62], yielding an uncertainty of 2 kJ/mol (3 kJ/mol for urea) at the main barrier and/or the main energy well.

3.3 Results

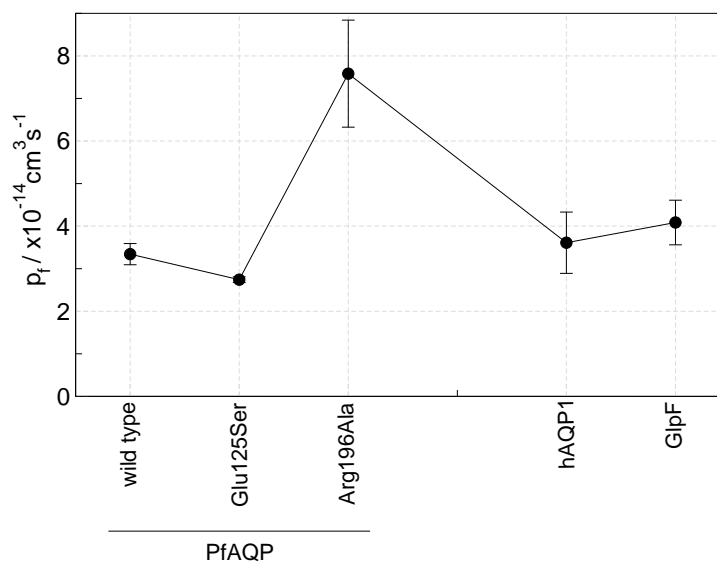


Figure 3.2. Single-molecule water permeability coefficients derived from equilibrium molecular dynamics simulations for the indicated proteins and mutants.

To quantify the water permeation through the different studied aqua(glycero)porins, the permeability coefficient, p_f , was computed from equilibrium molecular dynamics simulations (Fig. 3.2). The computed p_f value of the wild type PfAQP is $3.3 \pm 0.2 \times 10^{-14} \text{ cm}^3/\text{s}$. This value is nearly equal to the computed value for hAQP1 and slightly smaller than that of GlpF. In addition, the mutation of Glu125 (at the C loop, Fig. 3.1) into serine shows a reduction in the p_f of 18% compared to the wild type simulation. In contrast, the mutation of Arg196 (inside the narrow ar/R region of the pore, Fig. 3.1) into alanine reveals a substantial increase (by a factor of two) in the water permeability compared to the wild type simulation.

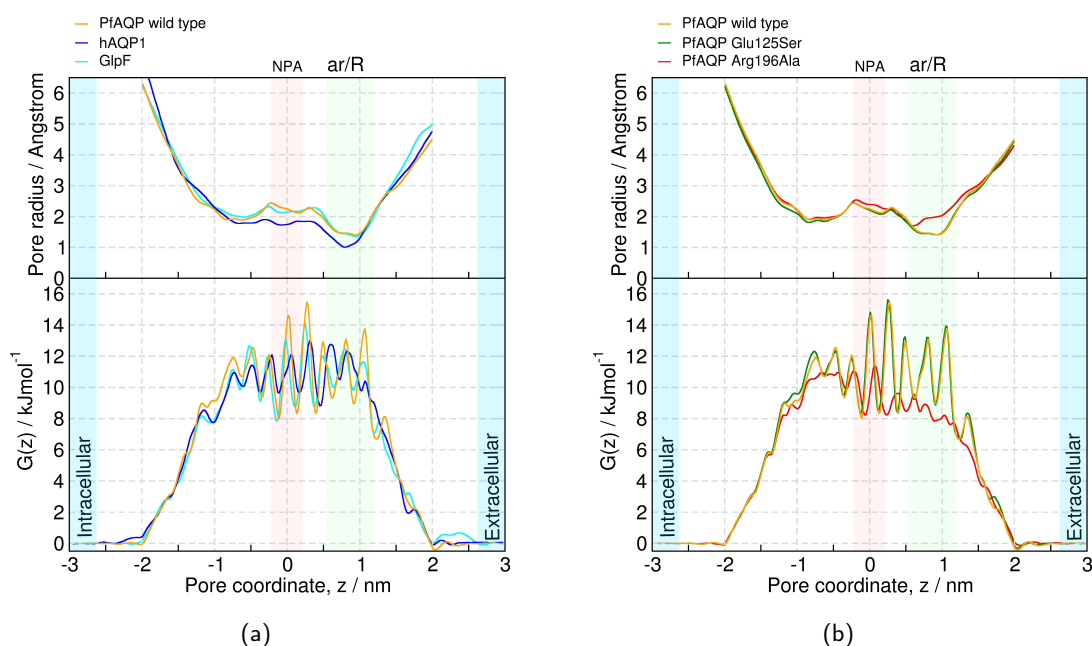


Figure 3.3. Pore radius profiles (upper panels) and potential of mean force for water permeation profiles (lower panels) derived from equilibrium molecular dynamics simulations. (a) Comparison between PfAQP (orange), hAQP1 (blue) and GlpF (cyan). (b) Effect of the point mutations Glu125Ser (green) and Arg196Ala (red).

To identify the rate limiting regions in the pore that determine the water permeation characteristics observed above, the averaged pore radius and the PMF for water were calculated (Fig. 3.3). In PfAQP, hAQP1 and GlpF, the pore geometry remains essentially unchanged compared with the hole profiles calculated from the X-ray structures [55], implying a relatively rigid pore. In PfAQP (orange line) and GlpF (cyan line) the pore profiles were found to be remarkably similar, narrowing in the ar/R region down to 1.4 Å radius, and wider than the profile in hAQP1 (blue line). In addition, the point mutation Glu125Ser (green line) does not modify the pore geometry, whereas the Arg196Ala mutation (red line) induces a widening of the pore to 1.7 Å at the ar/R region.

The PMF for water permeation is also illustrated in Fig. 3.3. In PfAQP (orange line), the PMF displays two major barriers of 14.6 kJ/mol and 15.4 kJ/mol near the NPA region which is not the most constricted region, and a third smaller barrier of 13.8 kJ/mol at the end of the ar/R region. In hAQP1 (blue line) and GlpF (cyan line), the PMFs also have a main free energy barrier at the NPA region, but they are smaller than the barrier in pAQP by 2.4 kJ/mol and 1.3 kJ/mol, respectively. Furthermore, in PfAQP, the PMF displays eight local minima inside the channel, and water molecules are predominantly found at the positions of these minima. Consequently, the effective water-water distance along the pore, computed as the average

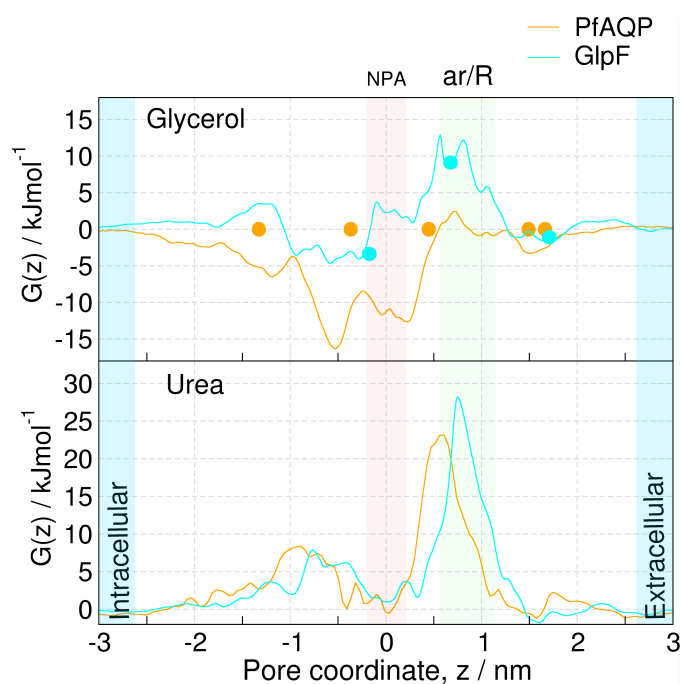


Figure 3.4. PMFs for glycerol and urea permeation through PfAQP (orange) compared with previous molecular dynamics studies for GlpF [58] (cyan). The dots correspond to the glycerol positions in the PfAQP [55] (orange dots) and GlpF [50] (cyan dots) crystal structures.

distance between minima, is $2.6 \pm 0.3 \text{ \AA}$. At the same region, the PMF for hAQP1 (GlpF) shows seven (eight) minima separated by a water-water distance of $2.9 \pm 0.6 \text{ \AA}$ ($2.8 \pm 0.6 \text{ \AA}$). In hAQP1, the water-water distance increment is mainly observed between the NPA and the ar/R region. Additionally, the Glu125Ser mutant (green line) does not reveal significant changes in the PMF profile, whereas the Arg196Ala mutant (red line) appeared to have a complete flattened profile at the ar/R region with a corresponding decrease in the free energy barrier to a value less than 11 kJ/mol.

Finally, Fig. 3.4 depicts the PMFs for glycerol and urea, derived from umbrella sampling simulations, for PfAQP (orange line), and compared with previous molecular dynamics studies for GlpF [58] (cyan line). In PfAQP, the free energy barrier at the ar/R region for glycerol is substantially smaller than the existing for urea (approximately 21 kJ/mol smaller). Moreover, for both solutes, glycerol and urea, the free energy barrier at the ar/R region decreases in PfAQP compared to that observed in GlpF. In addition, the PMF for glycerol reveals new features in PfAQP that are not observed in GlpF: first, there is a potential well at the intracellular vestibule ($z < -1.0 \text{ nm}$) instead of a free energy barrier. Second, the potential wells at an intermediate region between the intracellular vestibule and the NPA motifs are deeper in PfAQP than in GlpF. Contrary to glycerol, the PMFs for urea in PfAQP and GlpF were found to be similar between the intracellular vestibule and the NPA motifs. Furthermore, for both PfAQP and GlpF, glycerol crystallographic positions (indicated with dots) agree favorably with the potential well positions observed in the simulations.

3.4 Discussion

Here we present a computational study to characterize at the molecular level the permeation of water, glycerol and urea through the *plasmodium falciparum* aquaglyceroporin (PfAQP). Initially, the water permeation through PfAQP was quantified by means of the single molecule osmotic permeability coefficient, p_f . Subsequently, the pore geometry and the energetics for water, glycerol and urea transport through PfAQP was analyzed by computing radius and potential of mean force profiles.

3.4.1 Quantifying the water transport activity

Let us first consider the absolute p_f values predicted in our simulations in light of permeation measurements and previous computational studies. First, our computed values for the reference (aquaporin) hAQP1 and (aquaglyceroporin) GlpF (Fig. 3.2) are close to the experimental values, spanning between $4.6 \times 10^{-14} \text{ cm}^3/\text{s}$ to $11.7 \times 10^{-14} \text{ cm}^3/\text{s}$ for hAQP1 [31, 32, 34, 143] and $0.7 \times 10^{-14} \text{ cm}^3/\text{s}$ for GlpF [144]. Moreover, our predicted p_f for PfAQP is also close to these measurements. It should be noted that there are no single-molecule p_f measurements for PfAQP reported so far to compare directly. Second, our derived p_f values are smaller than the values predicted in previous computational studies on hAQP1 using the GROMOS force field and the SPC water model [4, 57] by a factor of two. Similarly, they are smaller than the values derived in studies on hAQP1 and GlpF using the CHARMM force field and the TIP3P water model [85, 145, 146] at least by a factor of three. These differences can be mainly attributed to the water diffusion constant. For the SPCE water model (used in this study) the diffusion constant is very close to the experimental value ($2.4 \times 10^{-5} \text{ cm}^2/\text{s}$). In contrast, for the SPC[147] and the TIP3P[140] water models (used in the mentioned previous computational studies) the diffusion constant is overestimated by a factor of 1.5 and 2, respectively.

Let us now to consider the relative differences in the the p_f for PfAQP compared to hAQP1 and GlpF. The computed p_f for PfAQP is close to the value predicted for hAQP1 (Fig. 3.2), suggesting that PfAQP conducts water at a similar rate as hAQP1. This result is in excellent agreement with experimental data from oocyte swelling assays [25], and supports therefore that PfAQP is a highly efficient water channel. Interestingly, the p_f for GlpF was observed to be higher than the values for PfAQP and hAQP1, suggesting that GlpF also conducts water at appreciable rates (comparable to hAQP1). A water permeability for GlpF has been observed in previous computational studies [4, 85, 145, 146], experimental assays with proteoliposomes [148] and reconstituted planar bilayers [144], but with the experiments

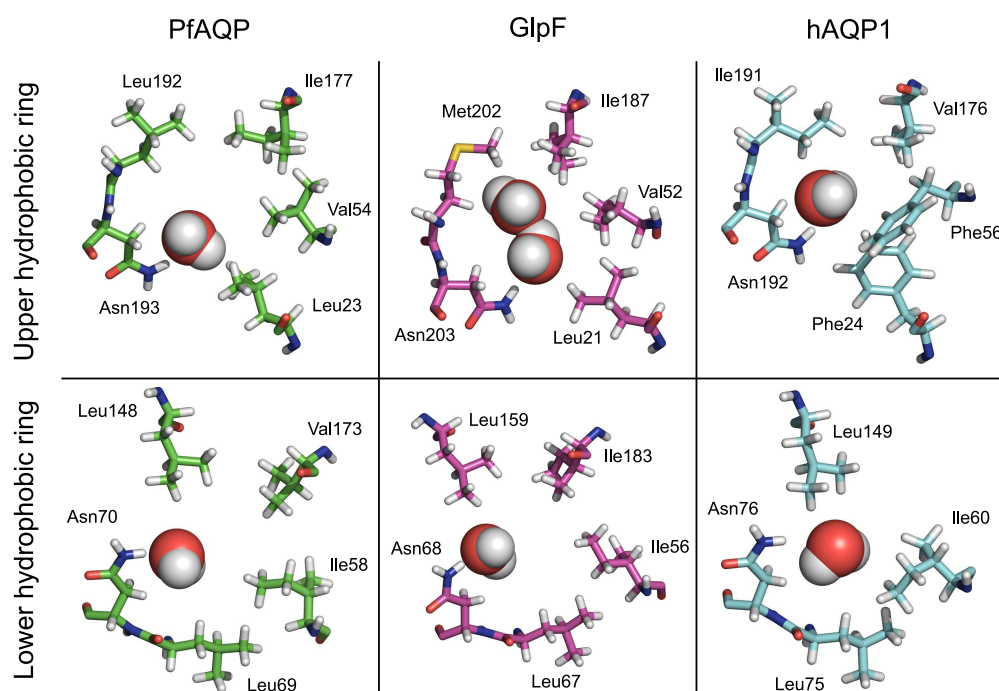


Figure 3.5. Top view of the two hydrophobic rings (upper and lower panels) located near the NPA motifs in the indicated proteins.

showing a lower permeation rate than the computational studies. The reason for this discrepancy remains unknown. A possible explanation would be simulation inaccuracies. However, given the excellent agreement between the simulation and the experiment for AQP1, and the consistent results for GlpF in several different and independent simulations, this appears unlikely. Another possible explanation could lie in a problematic estimation of the reconstitution efficiency, required to derive the copy number to estimate the experimental single-channel p_f .

3.4.2 The energetics of water permeation

Further calculations of the pore geometry and the PMF for water (Fig. 3.3) allowed us to identify the rate limiting regions that determine the water permeation properties presented above. The averaged radius profiles demonstrate that PfAQP and GlpF have practically identical pore geometries, wide enough to allow the passage of water and other solutes such as glycerol and urea. Accordingly, the narrower pore observed for hAQP1 constitutes one of the main factors for the exclusion of large solute molecules in this aquaporin. The average pore geometry from the simulations is remarkably similar to that from the initial crystal structures [42, 50, 55] (all compared in reference [55]), indicating a rigid channel.

The pore geometry itself is not sufficient to explain the water transport rates but the full

energetics must be taken into account. The major rate limiting region (highest barriers in the PMF) for PfAQP is observed at the NPA region, located exactly at the same place as for hAQP1 and GlpF (Fig. 3.3). In all three cases, two rings of hydrophobic residues sit there (Fig. 3.5), and they constitute the main barriers for water passage. This is in perfect agreement with previous simulations of hAQP1 and GlpF [4, 51], and confirms that PfAQP has similar water regulation mechanisms as other members of the family of aquaglyceroporins. Furthermore, in PfAQP, the Leu192 residue may force the permeating water molecules to interact more strongly with the Asn residue than in GlpF (where this Leu192 is replaced by Met202, a less hydrophobic residue, Fig. 3.5), leading to the observed higher free energy barrier, for PfAQP as compared to GlpF. Surprisingly, hAQP1 has smaller free energy barriers than PfAQP and GlpF in this critical region, despite the fact that the protein-water interactions are much stronger due to a narrower pore and the presence of the larger Phe24 and Phe56 residues in this region (Fig. 3.5)[4].

Finally, from the free energy barriers, it could be expected that PfAQP, GlpF and hAQP1 conduct water at rates in the following order: PfAQP < GlpF < hAQP1. Our p_f calculations predict the lowest rate for PfAQP. However, hAQP1 did not show the highest rate but GlpF did. This apparent contradiction between the p_f and the PMF for hAQP1 and GlpF may be due to the fact that a permeation event contributing to the p_f (the effective translocation of one water molecule from one aqueous medium to the other) implies the collective motion of all molecules inside the channel [83, 84], and therefore multiple free energy barriers and not only the highest should be taken into account when relating permeation rates to permeation energetics.

3.4.3 The effect of point mutations

We investigated the role of Glu125 and Arg196 on the water permeation of PfAQP. On one hand, the Glu125Ser mutant shows neither a substantial reduction in the p_f nor an increment in the free energy barriers compared to the wild type simulation (Fig. 3.2 and Fig. 3.3(b)), contrary to what previous mutation experiments demonstrated[134]. The reduction in the water permeability due to the Glu125Ser mutation was attributed to the destabilization of the C loop and further disruption of the hydrogen bonds between Arg196 and Trp124 [55, 134]. However, the conformational change of the C loop may occur in a longer time scale than the simulated time (100 ns), impeding to be detected in our equilibrium simulations. In fact, the C loop was observed to be rigidly anchored to the extracellular vestibule (with an average backbone rmsd of 0.97 Å with a standard deviation of 0.26 Å) in the wild type simulation, and just slightly more flexible (with an average backbone rmsd of 1.06 Å and a

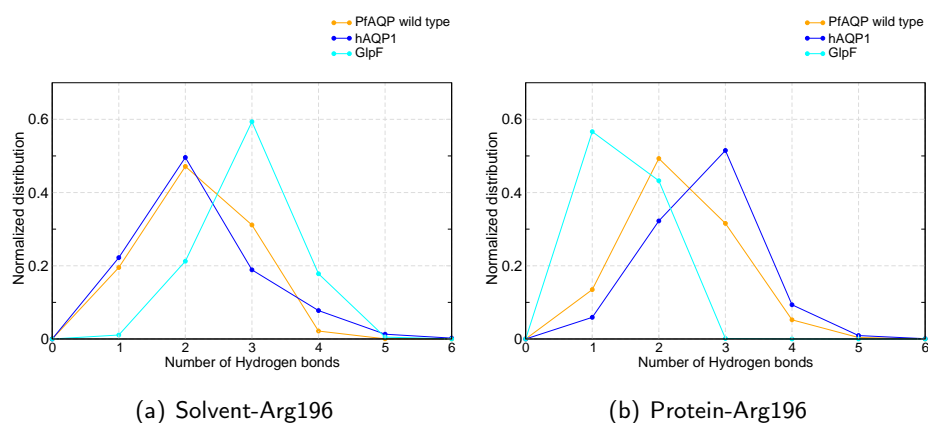


Figure 3.6. Water-Arg196 (a) and protein-Arg196 (b) hydrogen bond distribution in PfAQP (orange), hAQP1 (blue) and GlpF (cyan).

standard deviation of 0.36 Å) in the Glu125Ser simulation.

On the other hand, the Arg196Ala mutant reveals a widening of the pore at the Ar/R region and a corresponding reduction in the free energy barrier to water permeation, leading to a significant increase in the water permeability (Fig. 3.2 and Fig. 3.3(b)). This result confirms that Arg196 is a crucial residue for the water conduction through PfAQP.

3.4.4 Permeation of other solutes

We studied the energetics of glycerol and urea conduction (Fig. 3.4). Let us first analyze the free energy barriers at the ar/R region for the different simulated solutes. Our simulations predict that PfAQP conducts urea at a lower rate than glycerol which is in excellent agreement with oocyte assays [25]. In PfAQP, like in other aquaglyceroporins, glycerol molecules arrange at the ar/R in such a way that its hydroxyl groups can interact with Arg196, replacing water-Arg196 hydrogen bonds, and its apolar backbone orients towards the hydrophobic face (Trp50 and Phe190 residues) [50, 55]. In contrast, urea molecules cannot find this compensatory effect (dictated by the amphiphilic nature of the pore), because increasing the carbonyl-oxygen interactions with Arg196 would lead to unfavourable orientations of the amine groups facing the hydrophobic residues, and therefore would increase the free energy barrier as indeed observed in our simulations.

Focussing at the differences between PfAQP and GlpF, it is intriguing to note that PfAQP has a reduced free energy barrier for both glycerol and urea compared to GlpF at the ar/R region. This suggests that PfAQP is a more efficient glycerol and urea channel than GlpF, despite the fact that both have identical residues in this critical region. A possible explanation for this

would be that fewer water-arginine hydrogen bonds need to be replaced by the permeating solute molecule in the case of PfAQP compared to GlpF, and therefore the energetic cost to replace water with a solute at that position is reduced. In fact, Arg196 in PfAQP is forming three hydrogen bonds with the surrounding residues Ser200 and Trp124, whereas Arg206 in GlpF is only forming one with Phe135, as pointed out by Newby *et al.* [55]. Indeed, the histograms of the number of hydrogen bonds between the solvent and the arginine residue computed in the equilibrium simulations (Fig. 3.6(a)) show on average fewer hydrogen bonds for PfAQP than for GlpF, and accordingly, more hydrogen bonds between the protein and the arginine residue for PfAQP than for GlpF (Fig. 3.6(b)).

In addition, the glycerol PMF for PfAQP reveals two features not observed in GlpF: a potential well at the intracellular vestibule and a putative binding site inside the channel near the NPA region. This binding site is expected to affect the water permeability, particularly, at high glycerol concentrations. It is therefore highly interesting to see if this glycerol-dependent water permeability can be observed experimentally.

3.5 Conclusions

Here we present the first molecular dynamics study to characterize at the molecular level the permeation of water and other solutes through the *Plasmodium falciparum* malaria aquaglyceroporin (PfAQP). Our simulations confirm that PfAQP is a highly efficient water channel, that is able to conduct water at single-molecule permeability rates comparable to the rates for hAQP1 and GlpF. Furthermore, we identified the hydrophobic regions near the NPA motif as the main water rate limiting barriers, confirming therefore that PfAQP has a similar water regulation mechanism as other members of the family of aquaglyceroporins. Our simulations also support that GlpF is a highly efficient water channel, with a water permeability higher than that for hAQP1, as previously observed in several different and independent simulation studies.

We demonstrate that Arg196 located at the ar/R selectivity filter plays a crucial role regulating the permeation of water and other solutes such as glycerol or urea, as was previously hypothesized from functional and crystallographic studies. Our simulation results are consistent with a general permeation mechanism for aquaglyceroporins in which water is rate-limited by hydrophobic interactions in the NPA region. Other solutes, such as glycerol or urea, must accommodate in the ar/R selectivity filter, replacing water-arg196 hydrogen bonds interactions and facing the hydrophobic Trp50 and Phe190 residues, a process that is favorable for glycerol but less favorable for urea. In addition, in light of this mechanism, our simulations

suggest that PfAQP is able to conduct both glycerol and urea at higher permeabilities than GlpF does, because fewer water-Arg196 hydrogen bonds need to be replaced by a permeating solute molecule in the case of PfAQP compared to GlpF, a prediction to be further investigated and validated in experimental studies.

The effect on the water permeability due to the mutation of Glu125 (located at the C loop) to serine was found to be less severe in our simulations than observed experimentally. This could be attributed to the high stability of the C loop, rigidly anchored to the extracellular vestibule, that would slow C loop motions to time scales longer than the simulated time due to the mentioned mutation.

This study is expected to guide further computational and experimental studies in the search of putative blockers of PfAQP, understanding of their mechanism of action, that can hopefully be used to interrupt crucial physiological processes of the malaria parasite such as the water regulation and glycerol uptake.

The gating mechanism of aquaporins

Aquaporins play a central role for the water homeostasis of cells and its regulation is therefore a process of crucial importance. Regulation can occur either during trafficking or after translocation, i.e. when aquaporins are already inserted in the membrane. In the latter context, channel gating has emerged as an efficient regulatory mechanism, facilitating a rapid response to external stimuli when other regulatory mechanisms, such as transcriptional regulation or trafficking, are too slow [54].

Aquaporins have been proposed to be regulated by diverse mechanisms such as phosphorylation of serine residues, changes in divalent cation concentration, pH changes (these three reviewed in [149]), or applied membrane voltage [87].

In this chapter we employ molecular dynamics simulations to address the question on the molecular determinants triggering gating transitions of three different aquaporins: the yeast aquaporin (Aqy1), the spinach plant aquaporin (SoPIP2;1), and the human aquaporin 4 (hAQP4).

In section 4.1, we investigate the hypothesis that Aqy1 may be gated by either phosphorylation of a serine residue or membrane-mediated mechanical stress. Subsequently, in section 4.2, for SoPIP2;1, we examine how phosphorylation of two serine residues and protonation of a histidine residue, the two key components of the current proposed mechanism for plant aquaporins, affect the water permeation and the structure of the postulated gate of the pore. Finally, in section 4.3, we explore if hAQP4 may be voltage gated by a similar mechanism as the one proposed in a previous simulation study for hAQP1 [87].

4.1 The yeast aquaporin: gating mediated by phosphorylation and mechanosensing

4.1.1 Introduction

An unusual characteristic of yeast aquaporins is that they frequently contain an extended N-terminus of unknown function. The X-ray structure of the yeast aquaporin (Aqy1) from *Pichia pastoris* at 1.15 Å resolution was recently determined [54]. It revealed that the water channel is closed by the N-terminus, which arranges as a tightly wound helical bundle, with Tyr31 forming H-bond interactions to a water molecule within the pore and thereby occluding the channel entrance (figure 4.1(d)). Nevertheless, functional assays show that Aqy1 has appreciable water transport activity which aids survival during rapid freezing of *P. pastoris*. These findings establish that Aqy1 is a gated water channel. Here we address the question of a putative gating mechanism for Aqy1 by using molecular dynamics simulations. We investigate the hypothesis that the Aqy1 may be regulated by both phosphorylation of a serine residue or membrane-mediated mechanical stress.

4.1.2 Methods

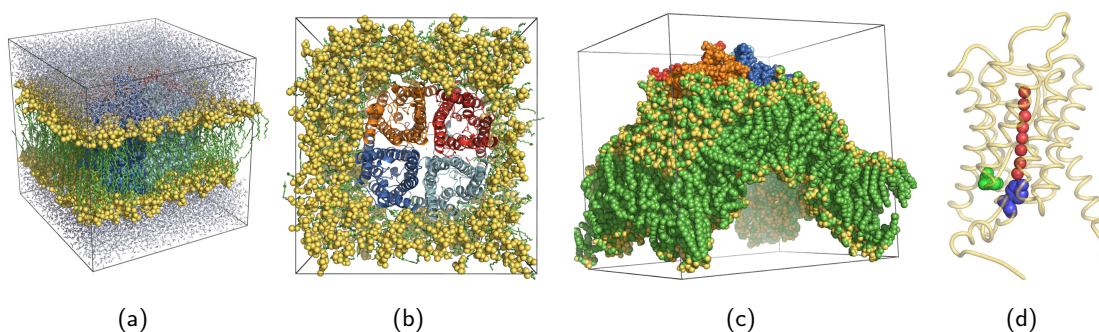


Figure 4.1. Molecular dynamics simulations of Aqy1. (a)-(c) Simulation boxes showing the tetramer (blue, orange, red, and cyan), fully embedded in a POPE lipid bilayer (yellow head groups and green tails) and solvated by water (blue, white). (c) Snapshot of the simulation inducing a bending of the membrane. (d) Aqy1 monomer showing the Ser107 (green) and the Tyr31 (blue) residues.

Molecular dynamics simulations were carried out starting with the Aqy1 tetramer in a fully solvated palmitoylphosphatidylethanolamine (POPE) lipid bilayer (see figure 4.1). Five different simulations were carried out: first, control simulations under equilibrium conditions, without exerting external forces over the membrane (I); second, mutating Ser107 into Asp, mimicking a putative phosphorylated state (II); third, mutating Tyr31 into Ala (III); fourth, inducing a surface tension onto the membrane (IV), and fifth, bending the membrane towards

the cytoplasmic side (V). Mutations for the simulations II and III were carried out with the WHAT IF software [136].

The simulation boxes contain the protein tetramer, 268 POPE lipids (361 for the bent membrane in simulation V) and 18580 (30842 for the bent membrane in simulation V) SPC water molecules [147]. Crystallographic water molecules were kept in the structures, and chloride ions were added to neutralize the simulation systems. The OPLS all-atom force field [101, 102] was used for the protein, and lipid parameters were taken from Berger *et al.* [138]. The simulations were carried out by using the GROMACS simulation software [98, 99]. Electrostatic interactions were calculated with the particle-mesh Ewald method [114, 115]. Short-range repulsive and attractive interactions were described by a Lennard-Jones potential, which was cut off at 1.0 nm. The Settle algorithm [109] was used to constrain bond lengths and angles of water molecules, and Lincs [110] was used to constrain all other bond lengths, allowing a time step of 2 fs. The temperature was kept constant by weakly coupling the protein, lipids and water molecules separately to a heat bath at 300 K [116] with a coupling constant $t = 0.01$ ps. The pressure was kept constant by weakly coupling the system to a semiisotropic pressure bath at 1 bar with a coupling constant of $t = 1$ ps. To induce a surface tension in the simulation IV, the pressure in the direction parallel to the membrane surface (xy plane) was increased to 10 bar, and the pressure in the direction normal to the membrane was kept at 1 bar. To bend the membrane in the simulation V, an external force was exerted on lipids located at distances larger than 5 nm from the centre of the tetramer along the z coordinate, resulting in a net constant acceleration of 0.01 nm ps^{-2} . A compensation force was applied to the protein to prevent net acceleration of the simulation box. All simulations were equilibrated for 1 ns before production. During this time the coordinates of the protein were harmonically restrained, with a harmonic force constant of $1000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$. The simulation length in the simulations I to IV was 100 ns whereas in simulation V it was 10 ns.

Pore dimensions

Pore diameter profiles were obtained with the HOLE software [139], averaging over several snapshots taken each 50 ps in a time window given by $t_{\text{start}} - t_{\text{end}} = 95 \text{ ns} - 100 \text{ ns}$, $95 \text{ ns} - 100 \text{ ns}$, $50 \text{ ns} - 55 \text{ ns}$, $21 \text{ ns} - 25 \text{ ns}$ and $6 \text{ ns} - 10 \text{ ns}$, for the simulations I to V, respectively.

Potential of mean force for water

The potential of mean force for water was obtained following the procedure described in section 2.2.2. It was independently calculated for every monomer using the equation 2.24 [4], by computing the average (over the whole trajectory) of the number of water molecules

at the z position along the pore, $\langle n(z) \rangle$. The number of water molecules is confined to a cylinder that is aligned with the pore coordinate. Therefore, the entropy in the bulk regions is reduced, and consequently, the free energy inside the channel referred to the bulk would be underestimated. To relate $G(z)$ with the area of one aquaporin monomer and correct for such underestimations, a trapezoidal correction was applied in the entrance and exit regions (see equation 2.25). In consequence, the final PMF refers to a density of one channel per membrane area occupied by an aquaporin monomer [58]. The computed correction was 3.0 kJ/mol. Prior to the analysis the protein monomers were superimposed to a reference structure. The effective PMF, $G_{\text{eff}}(Z)$, was also computed by combining the four monomer $G_i(z)$ by using the equation 2.26.

Phe92 was observed to occasionally flip and block the pore in between the NPA and the ar/R region. These motions increased the energy barrier for water permeation in simulation II. The significance of this motion is unclear, and it was found to be independent of the opening events taking place near Tyr31.

Principal component analysis

A principal component analysis, consisting of the calculation and diagonalization of the covariance matrix of the atom coordinates, was carried out to detect relevant collective motions associated with the gating transitions. It was performed considering all the coordinates of the backbone atoms of the lower part of helices four, five and six, and loop D (see figure 4.3), accumulated over the whole trajectory [130], was carried out using the GROMACS package tools (see section 2.2.3).

Essential dynamics simulations

To investigate the correlation between the conformational changes taking place in the gate of the pore and the collective coordinate represented by the first eigenvector (obtained in the PCA analysis), a series of 10 ns essential dynamics simulations were carried out [130]. The backbone atoms of lower parts of helices four, five and six, and the loop D were forced to move along the principal eigenvector found in the PCA analysis (figure 4.3), with a constant driving velocity of 1 nm/ns starting from the closed conformation. After 0.8 ns, 0.9 ns, 1.0 ns and 1.1 ns the driving velocity was set to zero nm/ns, and the atoms were constrained to maintain the projection value constant until the end of the simulation. These simulations were also carried out with the GROMACS software tools.

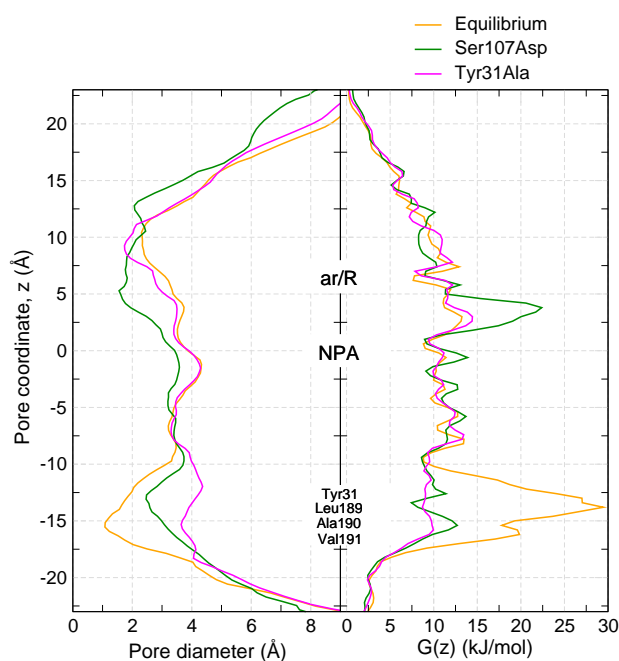


Figure 4.2. Pore diameter and free energy profiles derived from molecular dynamics simulations of Aqy1 (orange), Aqy1S107D (green), and Aqy1Y31A (magenta).

4.1.3 Results

Tyrosine residue blocks the channel

The X-ray structure of Aqy1 at a resolution of 1.15 Å was recently determined [54]. It revealed that the water channel is closed by the N-terminus, which arranges as a tightly wound helical bundle, with Tyr31 forming H-bond interactions to a water molecule within the pore and thereby occluding the channel entrance (figure 4.1(d)). To further investigate the stability of this closed conformation and the role of Tyr31, simulations of the wild type Aqy1 and the Tyr31Ala mutant were carried out. In the simulation of the wild type Aqy1 (orange line in the figure 4.2), the pore diameter profile reveals how the water channel narrows to 0.8 Å in diameter near Tyr31, which is too small to allow the passage of water. Consequently, there is an energetic barrier to water permeation of approximately 30 kJ/mol in this region, confirming the closed nature of the channel. When Tyr31 is substituted with an alanine (magenta line in the figure 4.2) the pore widens to a diameter larger than 2 Å, which reduces the free energy barrier to water permeation to less than 13 kJ/mol, and water molecules enter the channel and fill the space left by Tyr31. These results are in qualitative agreement with functional assays that showed a 6-fold increase in the water transport activity when Tyr31 of Aqy1 is substituted by an alanine [54].

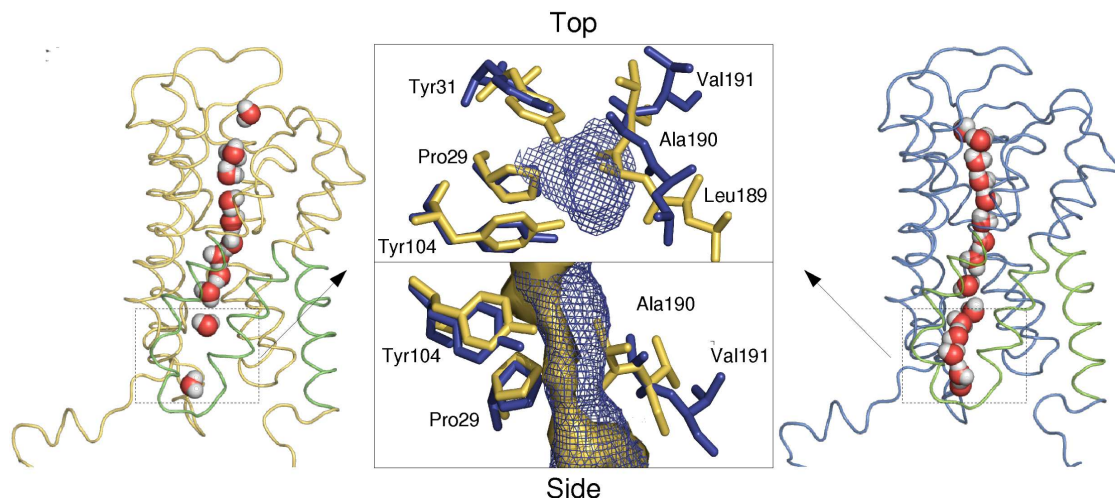


Figure 4.3. Snapshots of the closed (left) and the open (right) conformations showing how a water column is established in a similar way upon opening, during either simulations of the mutant *Aqy1S107D*, or increasing the lateral pressure to 10 bar, or bending the membrane towards the cytoplasmic side (the region highlighted in green was considered for the principal component analysis mentioned in the text). Top and side views of the region indicated by the dashed squares are shown in the middle panel: The pore surface is shown for both the closed (yellow) and the open (blue) conformations. The channel widens after a local rearrangement of residues *Leu189*, *Ala190*, and *Val191*, and this motion is observed for all three simulations.

Putative phosphorylation site triggers channel opening

The structure of Aqy1 also revealed that Ser107 lies within a consensus phosphorylation site situated near the pore channel and is involved in an important network of hydrogen bonds involving Tyr31 [54]. We investigated the effect of the substitution of Ser107 by aspartate, mimicking a putative phosphorylation event. Molecular dynamics simulations of the S107D Aqy1 mutant show a widening of the pore near Tyr31, which increases to a diameter larger than 2 Å, and a corresponding reduction in the free energy barrier to water permeation to less than 13 kJ/mol (green line in the figure 4.2). In these simulations water molecules establish a single-file water column between Pro29, Tyr31, Tyr104, Leu189, Ala190, and Val191, after a local rearrangement of the latter three residues, which are located in the lower part of helix 4 towards loop D (figure 4.3). In contrast, simulations of the Y31A Aqy1 mutant allowed water molecules to enter the channel, filling the space left by Tyr31. Accordingly, water transport assays in spheroplasts reveal a significant increase in water transport activity when Ser107 of Aqy1 was substituted by aspartate [54]. Thus both molecular dynamics simulations and functional data support the suggestion that Ser107 is a putative phosphorylation site, and that it can induce an opening of the pore upon phosphorylation.

Regulation by Mechanosensitivity

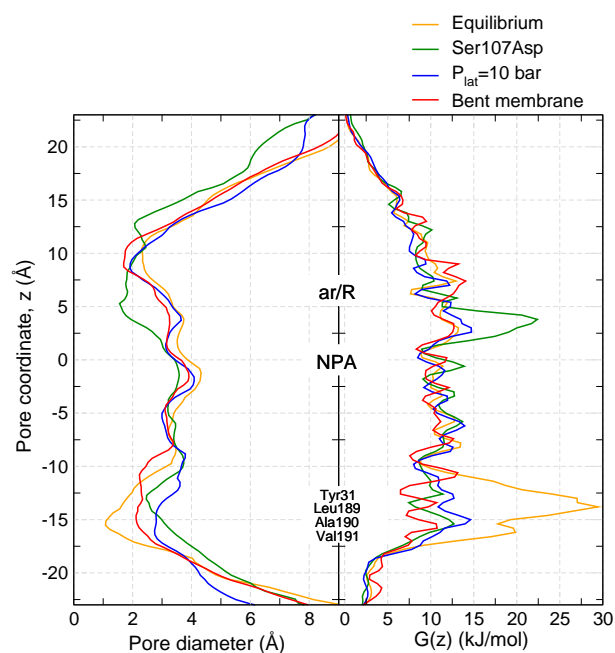


Figure 4.4. Pore diameter profiles (left), showing a widening of the channel near the Tyr31 region for simulations with the mutated Aqy1S107D (green), with an increased lateral pressure (blue), or with a bent membrane (red), compared with the control simulation (yellow). Free energy profiles (right), showing a substantial reduction in the energetic barrier for water permeation in the monomers that opened during the simulations.

Functional assays revealed that Aqy1 have higher water-transport activity when purified and reconstituted into proteoliposomes than it has in its native membrane [54]. A clue hinting at an explanation for this observation is that Aqy1 may be a mechanosensitive channel. Aqy1 would sense the mechanical stress induced by the membrane. Accordingly, it would be more active in proteoliposomes than in its native membrane, because these proteoliposome vesicles are highly curved (120-130 nm in diameter) compared to spheroplasts native membranes (with a diameter of 1-5 mm) [54].

To test this hypothesis, nonequilibrium molecular dynamics simulations of Aqy1 were performed in a solvated lipid bilayer being subject to external mechanical stress, either by increasing the lateral pressure up to 10 bar, or by bending the membrane towards the cytoplasmic side (figure 4.1(c)). Spontaneous opening events of one monomer in both simulations were observed (figure 4.3) with the pore diameter near Tyr31 widening from 0.8 Å in the crystal conformation to values larger than 2 Å in both simulations (figure 4.4). In addition, the energetic barrier for the water permeation dropped substantially compared to the control simulations in which no opening was observed.

4.1.4 Discussion and conclusion

Here we address the question of a putative gating mechanism for Aqy1 by using molecular dynamics simulations. We investigate the hypothesis that the Aqy1 may be regulated by either

phosphorylation of a serine residue (Ser107) or membrane-mediated mechanical stress. Our simulations confirm that under equilibrium conditions the Aqy1 channel is blocked by a tyrosine residue (Tyr31) in the N-terminus (figure 4.3). This is in excellent agreement with functional assays and the high resolution X-ray structure of Aqy1 [54]. Both molecular dynamics simulations and functional data support the suggestion that Ser107 is a putative phosphorylation site, and that it induces an opening of the pore upon phosphorylation (figures 4.2 and 4.3). However, this does not explain why Aqy1 has a high water-transport activity when purified and reconstituted into proteoliposomes than it has in its native membrane. Thus mechanosensitive gating emerges as a plausible mechanism. This possibility is also supported by our molecular dynamics simulations, which show how Aqy1 can be regulated by both the surface tension and membrane curvature (figures 4.3 and 4.4).

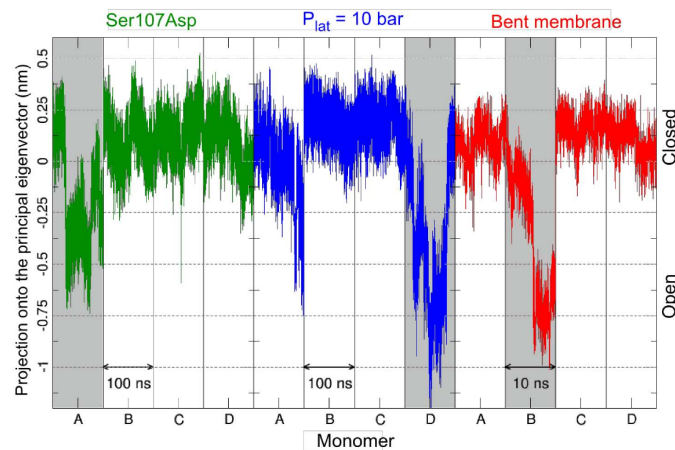


Figure 4.5. Projections of nonequilibrium trajectories onto the principal eigenvector obtained by principal component analysis. Transitions are observed upon opening for the monomer A in the left panel (Aqy1S107D), D in the middle panel ($P_{lat} = 10$ bar), and B in the right panel (bent membrane).

An intriguing question for the mechanism of mechanosensitivity is how a mechanical signal is transmitted from the membrane to the gate of the channel in the cytoplasmic portion of helix 4, which is not in direct contact with the membrane. To address this question, a principal component analysis of the backbone atoms of the cytoplasmic halves of helices four, five, six, and loop D was carried out (region shown in green in figure 4.3). Projections of both membrane-mediated stress trajectories onto the principal eigenvector show an abrupt transition for the monomers in which an opening takes place (figure 4.5), suggesting that external forces triggering gating are transmitted from the lipid membrane to Leu189, Ala190, and Val191 via coupled movements of the helices four, five, and six, the latter being in direct contact with the membrane.

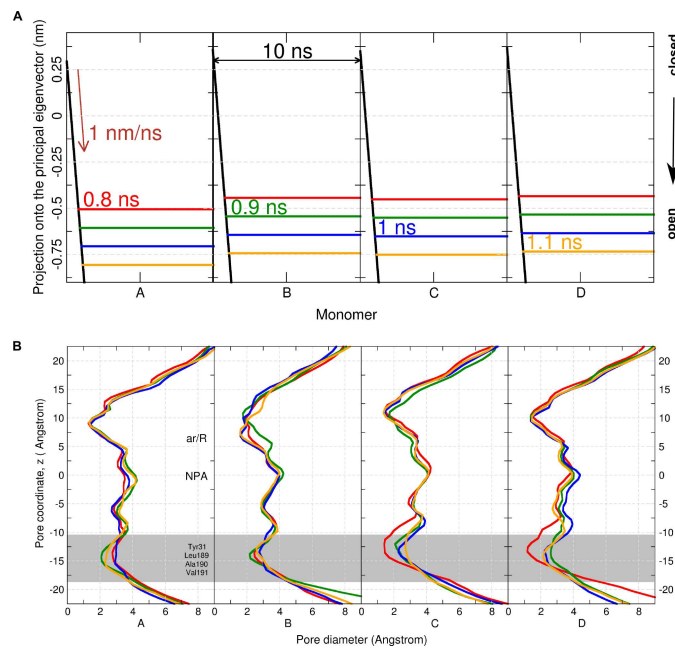


Figure 4.6. Essential dynamics simulations of Aqy1. (A) Projections of all trajectories onto the principal eigenvector found in the PCA analysis. Initially, for each monomer the driving velocity was 1 nm/ns (black lines). Then, after 0.8 ns (red), 0.9 ns (green), 1.0 ns (blue), and 1.1 ns (orange), the driving velocity was set to 0 nm/ns. (B) Averaged pore diameter profiles for each one of the four monomers, during the second part of the simulation (driving velocity equals 0 nm/ns), after 0.8 ns (red), 0.9 ns (green), 1.0 ns (blue), and 1.1 ns (orange). The pore widens near Tyr31 to values larger than 2 \AA for 14 of the 16 cases, indicating that the collective coordinate represented by the first eigenvector is indeed responsible of the gating conformational change, and that the opening motions taking place at the gate of the pore are reproducible separately in the four monomers.

To investigate the causal relation between this global conformational change and channel opening, the position along this principal eigenvector was artificially driven from the "closed" to the "open" conformation in an additional set of essential dynamics simulations. The pore widens near Tyr31 to values larger than 2 \AA for 14 of the 16 cases, indicating that the collective coordinate represented by the first eigenvector is indeed responsible of the gating conformational change, and that the opening motions taking place at the gate of the pore are reproducible separately in the four monomers (figure 4.6).

It is intriguing to consider how the two putative modes of regulation, by both phosphorylation of a serine residue or changes in the surface tension and curvature of the membrane, may relate to each other. Strikingly, the opening transition for the Aqy1S107D simulation, which mimics a phosphorylated S107, involves similar movements of residues Leu189, Ala190, and Val191 as in the simulations with the membrane being subject to external mechanical stress (figure 4.3). These similarities are revealed in a principal component analysis of the Aqy1S107D trajectory, where the projection onto the first eigenvector, upon opening, drops in a manner

similar to simulations with induced membrane stress (figure 4.5). These findings disclose that both phosphorylation and external membrane-mediated mechanical stress may induce a similar opening mechanism, and suggest how mechanical stimuli are transmitted from the membrane to the gate of the channel.

Regulation of Aqy1 by phosphorylation may also be exploited in physiological contexts, when less rapid changes in water transport activity are required. Hence, phosphorylation may fine-tune the water flux during normal conditions of growth, whereas mechanosensitive gating could provide a rapid pressure valve in response to unexpected shocks. Rapid freezing or thawing and sudden osmotic changes are frequently encountered by micro-organisms [150]. Thus the evolution of gated aquaporins and aquaglyceroporins would provide an economic solution to numerous stresses associated with rapidly changing environments, aiding the organism's quest to adapt and survive.

4.2 Is the spinach plant aquaporin gated by phosphorylation or histidine-protonation?

4.2.1 Introduction

In plants, aquaporins intervene in vital processes during growth and development, and help to cope with sudden environmental changes [151, 152]. Several studies (reviewed in [152]) have stressed on the key roles that aquaporins play in plants, such as water transport through the tissues, transpiration, tissue expansion, nutrient uptake, and survival during drought or flooding conditions.

A gating mechanism involving phosphorylation and pH changes was proposed for plant aquaporins [48, 153, 154]. The mechanism proposes that aquaporins in the plasma membrane (plasma membrane intrinsic proteins, PIPs) open and close to counteract large osmotic gradients between the cell interior and exterior [48]. According to the model, during drought stress conditions, the PIPs close in response to the dephosphorylation of two highly conserved serine residues located in the cytosolic loop B and the C-terminus, respectively. Accordingly, during flooding, the pH at the cytoplasmic medium decreases, and PIPs close in response to the protonation of a conserved histidine located at the cytosolic loop D.

The X-ray structures of the spinach aquaporin SoPIP2;1 provided structural insights into the gating mechanism [48]. It revealed the aquaporin in two different conformations: closed and open (figure 4.7(b) and figure 2 in reference [48]). The main difference between them concerns the structure of loop D: in the closed conformation, it is near helix 1 with the Leu197 occluding the pore, whereas in the open conformation, it moves towards the C-terminus leaving the pore fully open. In addition, molecular dynamics simulations described the initial opening stages upon phosphorylation of S115 and S274 [48], and identified important residues at loop D and helix 1 that stabilizes the closed conformation [155]. Furthermore, functional studies have stressed on the role of the histidine at the loop D (H193 in SoPIP2;1) for the pH gating [154, 156].

Despite of all this evidence, some aspects of the proposed gating mechanism remain controversial. First, the water permeability did not substantially decrease (only up to 5 fold) when lowering the pH between 7 and 6, as it would be expected for a complete closure of the channels during flooding conditions. Second, mutational assays of the tobacco NtPIP2;1 and NtAQP1 suggested that dephosphorylation of the key serine residue at the C-terminus (S274 for SoPIP2;1) does not modify the pH water-permeability dependence [156], contrary to what the gating model predicts. Third, single-molecule water permeability measurements of SoPIP2;1 (W. Kukulski and A. Engel, Biozentrum, private communication, 2008) revealed

high water permeability rates for both the wild type protein and mutants mimicking a dephosphorylated state of S115 and S274. This suggests therefore that dephosphorylation of these two serine residues does not affect the water conductivity of the channel. Fourth, two additional X-ray structures of SoPIP2;1 [49], with S115 or S274 mutated by glutamic acid mimicking the phosphorylated state, were not obtained in an open conformation, as expected from the gating mechanism.

Here, we carry out equilibrium molecular dynamics simulations to address these issues and gain insights into the regulatory mechanism of SoPIP2;1. We investigate how phosphorylation of the two mentioned serine residues and the protonation of the histidine affects the single-molecule water permeability of the channel and the conformation of the loop D. We also analyze the impact of the deletion of the C-terminus on these two functional and structural observables. Finally, we explore the possibility that SoPIP2;1 might be gated by membrane-mediated mechanical stress.

4.2.2 Methods

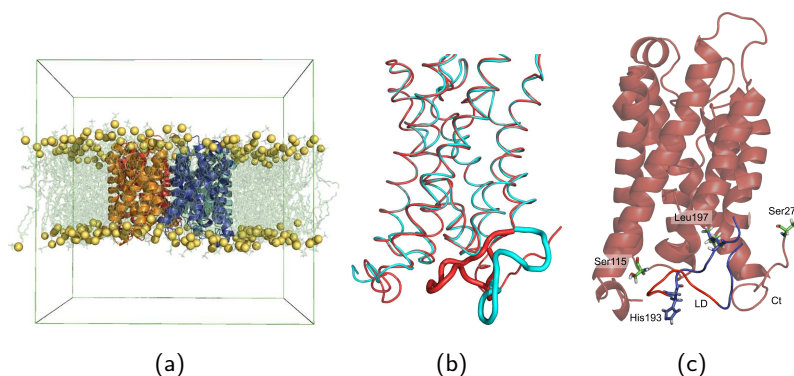


Figure 4.7. Molecular dynamics simulations of SoPIP2;1. (a) Simulation box showing the SoPIP2;1 tetramer (cartoon representation), embedded in a lipid bilayer (yellow head groups and green tails) and solvated by water (not shown). (b) Closed (red) and open (cyan) X-ray structures of the loop D [48]. (c) Monomer indicating the loop D (LD), the C-terminus (Ct) and the putative residues involved in gating transitions. The two ends of the loop D were considered for the PCA analysis (blue) as mentioned in the text.

Simulation details

Molecular dynamics simulations were carried out starting with the SoPIP2;1 aquaporin tetramer in a fully solvated lipid bilayer (figure 4.7). Several independent simulations were carried out to study how the conformation of loop D and ultimately the water permeation are affected by external stimuli such as phosphorylation of two serine residues, protonation of a histidine

residue, deletion of the C-terminus or increase of the lateral pressure: First, control simulations of the wild-type form were carried out. Second, simulations adding a phosphate group (PO_3) to Ser115, Ser274, or both. Third, mutants mimicking (non-)phosphorylated conditions where Ser115, S274, or both were replaced by (alanine) aspartate. Fourth, simulations protonating His193, mimicking low pH conditions. Fifth, simulations deleting the C-terminus. Sixth, simulations inducing a surface tension, increasing the lateral pressure (table 4.1 and figure 4.7(c)).

	Simulation	N	Length (ns)	Force field
1	Wild type	1	97	amber03
		3	100	OPLS
2	S115+ PO_3	1	76	amber03
3	S274+ PO_3	1	76	amber03
4	S274+ PO_3 + S274+ PO_3	1	97	amber03
5	S115+ PO_3 + S274+ PO_3 + H193 ⁺	1	97	amber03
6	S115D	3	100	OPLS
7	S274D	3	100	OPLS
8	S115D+S274D	3	100	OPLS
9	S115D+S274D+H193 ⁺	3	100	OPLS
10	S115A	1	100	OPLS
11	S274A	1	100	OPLS
12	S115A+S274A	1	100	OPLS
13	H193 ⁺	3	100	OPLS
14	Δ Ct-SoPIP2;1	1	97	OPLS
15	Increased lateral pressure P_z normal to the membrane $P_x = 50$ bar, $P_y = 1$ bar $P_x = 1$ bar, $P_y = 50$ bar $P_x = 50$ bar, $P_y = 50$ bar			
		1	100	OPLS
		1	100	OPLS
		1	100	OPLS

Table 4.1. SoPIP2;1 simulations.

Simulations labeled from one to 14 in table 4.1 were carried out starting from both the open and closed crystallographic states, and simulations labeled 15 only from the closed state. The initial protein atom coordinates in the closed state were taken from the Protein Data Bank (PDB ID code 1Z98 [48]). Since the X-ray structure in the open state (PDB ID code 2B5F [48]) has a low resolution (3.9 Å), the initial atom coordinates in the open state were generated by pulling the loop D from the closed to the open X-ray state, by applying an harmonic force ($K = 1000$ kJ/mol) in an MD simulation of 10 ns. During this simulation, the C-termini were allowed to move to prevent overlap with the loop D, whereas the rest of the protein was kept fixed. Crystallographic water molecules and ions were kept in the structures, and chloride

ions were added to neutralize the simulation systems. The amber03 [157] and OPLS all-atom [101, 102] force fields were used for the protein. In the simulations with amber03, the tetramer was embedded in a patch of DOPC lipids [158] solvated by TIP3P water molecules [140], and in the simulations with OPLS, it was merged in a patch of POPE lipids [138] solvated by TIP4P water molecules [140]. Phosphoserine parameters used in simulations 2 to 5 were taken from Homeyer et al. [159] and different mutations of the S115 and S274 were carried out by using the WHAT IF modeling software [136]. The simulations were carried out by using the GROMACS simulation software [98, 99]. Electrostatic interactions were calculated with the particle-mesh Ewald method [114, 115]. Short-range repulsive and attractive interactions were described by a Lennard-Jones potential, which was cut off at 1.0 nm. The Settle algorithm [109] was used to constrain bond lengths and angles of water molecules, and Lincs [110] was used to constrain all other bond lengths, allowing a time step of 2 fs. The temperature was kept constant by weakly coupling the protein, lipids and water molecules separately to a heat bath at 300 K [116] with a coupling constant $t = 0.1$ ps. The pressure was kept constant by weakly coupling the system to a pressure bath at 1 bar with a coupling constant of $t = 1$ ps. All simulations were equilibrated for 1 ns before production. During this time the coordinates of the protein were harmonically restrained, with a harmonic force constant of $1000 \text{ kJmol}^{-1}\text{nm}^{-2}$. The simulation length of each simulation is shown in the table 4.1 for a total of $\sim 3.2 \mu\text{s}$.

Water permeability coefficients

The p_f was independently calculated for every monomer, based on the collective diffusion model [84] described in section 2.2.1. Water molecule displacements, dz_i , were computed every 10 ps, within a cylindrical region (of length $L = 2.5$ nm and radius $r = 0.45$ nm) centered at the pore axis and spanning -1.5 nm down and 1.0 nm up from the center of mass of the aquaporin tetramer. $\langle n^2(t) \rangle$ was obtained by averaging over 350 to 500 time windows (depending of the simulation length) of 200 ps length each. Finally the p_f was obtained from the slope of the curve $\langle n^2(t) \rangle$ versus time. An effective p_f value was obtained by averaging the values of the four monomers and the error was estimated as the standard error.

PCA analysis

A principal component analysis was also carried out here to detect correlations between putative gating motions of the loop D and the applied external stimuli, mentioned above. It was performed by considering all the coordinates of the heavy atoms of the two ends of the loop D (blue region in figure 4.7(c)) over the whole trajectory [130]. It was carried out using the GROMACS package tools (see section 2.2.3).

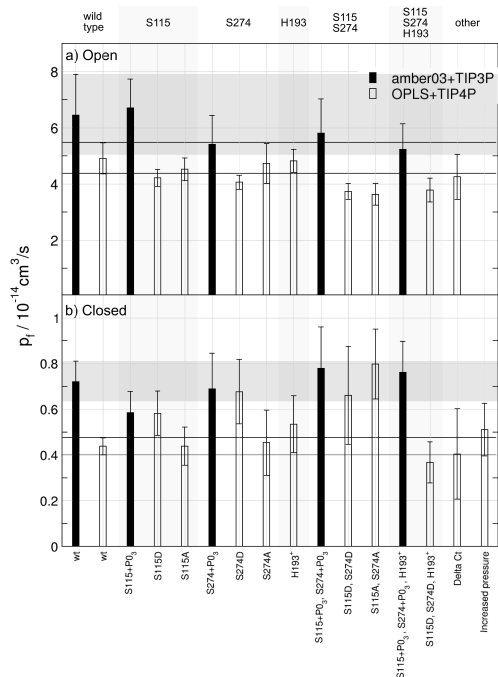


Figure 4.8. Water permeabilities derived from equilibrium MD simulations for the indicated protein and mutants. Simulations were carried out starting from both the open (upper panel) and the closed (lower panel) conformations. To guide the eye, the error in the wild type simulations is highlighted for both the amber03 and the OPLS force fields with the horizontal grey bar and the black lines, respectively.

4.2.3 Results

Water permeability calculations

The water permeability coefficient was computed from equilibrium MD simulations for the wild type protein and the above mentioned mutants (figure 4.8). The computed p_f values starting from the open state were approximately one order of magnitude larger than the values for the simulations starting from the closed state. This $p_f(open)/p_f(closed)$ ratio was independent of the used force field. The only observed difference is that the amber03+TIP3P force field yielded larger p_f values than the OPLS+TIP4P force field. This effect is attributed to the diffusion constant of the water molecules, as discussed previously in section 3.4.1. In the simulations starting from the open conformation using the OPLS+TIP4P force field, the p_f values were nearly two fold the value computed for hAQP1 using the same force field parameters ($p_f = 2.7 \pm 0.1 \times 10^{-14} \text{ cm}^3/\text{s}$). This is in agreement with single-molecule permeability measurements (W. Kukulski and A. Engel, Biozentrum, private communication, 2008) that showed that SoPIP2;1 conducts waters almost twice faster than hAQP1.

None of the perturbations, i.e. phosphorylation, protonation, deletion of the C-terminus or induced-membrane stress, strongly affected the average p_f values. Changes in the water permeability, in the order of the $p_f(open)/p_f(closed)$ ratio, were not observed. In the simulations starting from the open state, the average p_f values were found within the error of the values calculated in the wild type simulations, and the maximum reduction observed was 27 %, for the S115A+S274A double mutant. In addition, in the simulations starting from the closed state, a change up to two-fold was only observed for the S115A+S274A mutant.

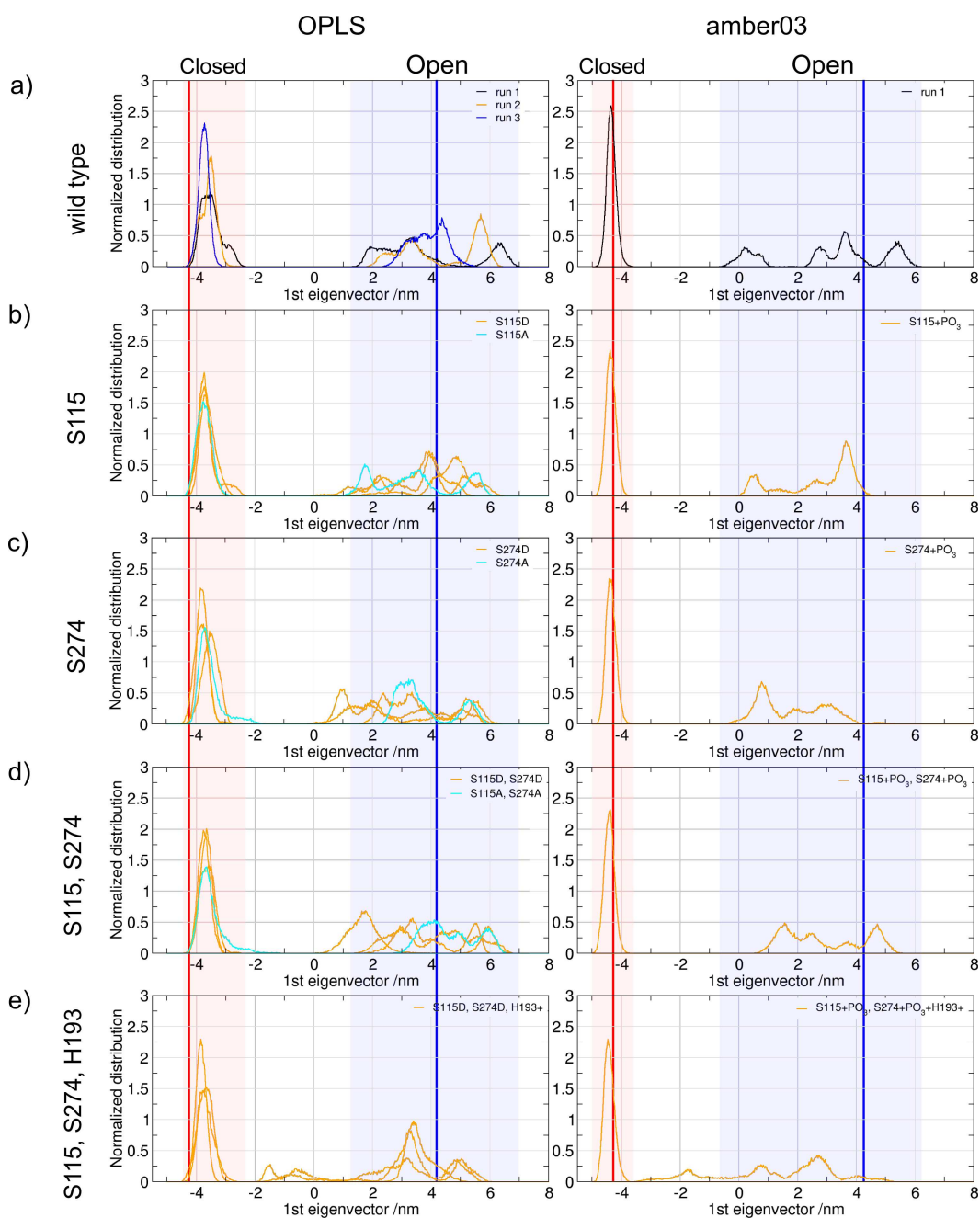


Figure 4.9. Normalized distributions of the positions along the difference PCA vector between the closed and open X-ray states, recovered from the projection of the MD trajectories onto this eigenvector. Simulations were carried out with the OPLS (left panels) and the amber03 (right panels) force fields, by mutating the indicated residues. Simulations starting from the closed (open) conformation yielded the distributions at the left (right) side of each panel and from the open conformation at the right side. The regions visited in the wild type simulations are highlighted with the red (closed state) and the blue (open state) bars for comparison.

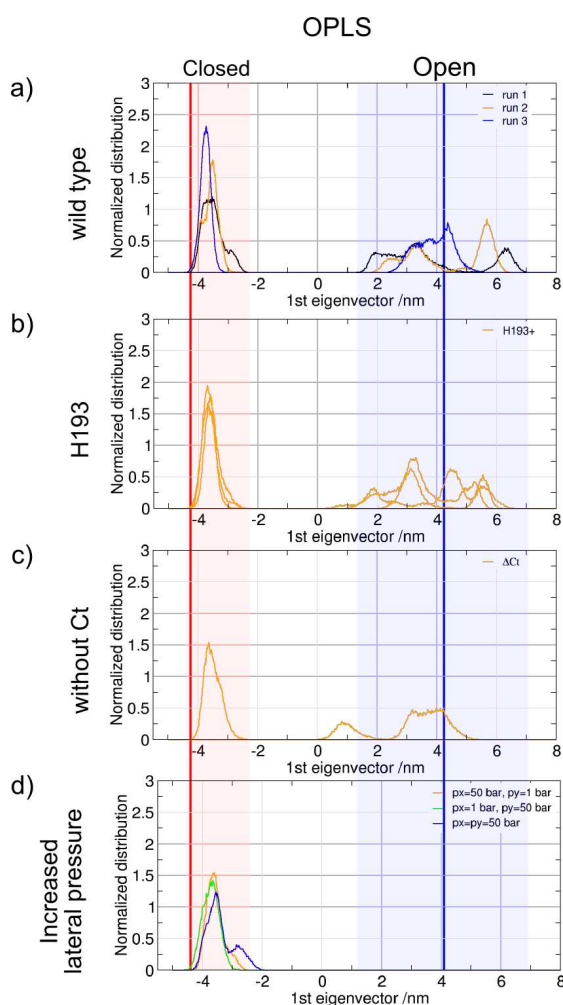


Figure 4.10. Normalized distributions of the positions along the difference PCA vector between the closed and open X-ray states, recovered from the projection of the MD trajectories onto this eigenvector. Simulations were carried out by mutating the indicated residues or inducing membrane-mediated mechanical stress. Simulations starting from the closed (open) conformation yielded the distributions at the left (right) side of each panel. The regions visited in the wild type simulations are highlighted with the red (closed state) and the blue (open state) bars for comparison.

Principal component analysis

The normalized distribution of the position along the difference PCA vector between the closed and open X-ray states was recovered, projecting the MD trajectories onto this eigenvector (figures 4.9 and 4.10). For the trajectories starting from the closed conformation, sharp narrow distributions (highlighted with red bars) near the X-ray position (vertical red line) were observed. Only transient opening transitions (right tails outside the red bar) were observed for the S274A and S115A+S274A mutants, and increasing lateral pressure. However, the structure of the loop D during these transitions remained practically unchanged. In contrast, for the trajectories starting from the open conformation, the distributions spread out more (region highlighted with blue bars), indicating a higher flexibility of the loop D in the open state than in the closed state. In these simulations, the distributions reached intermediate positions, corresponding to different degrees of closing, that involved relatively large conformational

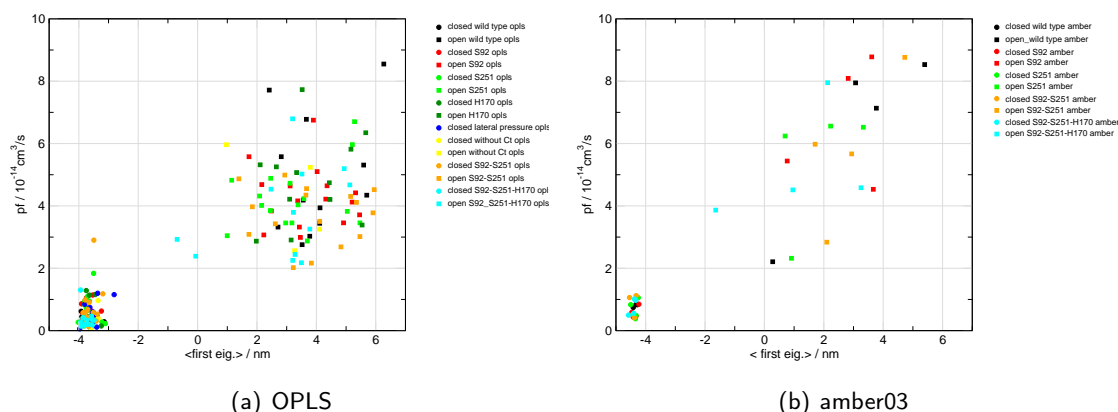


Figure 4.11. Correlation between the permeability coefficient, p_f , and the average position along the difference PCA eigenvector separating the closed and the open X-ray states, ($\langle \text{first eig.} \rangle$), for the indicated simulations.

changes of the loop D. In particular simulations mutating S115 and S274, and protonating H193 showed a significant widening of the distribution.

Correlation between the loop D conformation and the permeability coefficient

The correlation between the permeability coefficient of individual monomers and the average position along the vector connecting the closed and the open X-ray states was investigated (figure 4.11). On one hand, for the simulations starting from the closed state (circles), the average projections are concentrated near -4 nm and displayed significantly low p_f values compared to the values of the simulations starting from the open state (squares). On the other hand, in the simulations starting from the open state, the wide distribution along the difference vector, corresponding to different degrees of closing, is directly correlated to the p_f . However, none of the investigated externally applied stimuli, i.e. phosphorylation, protonation, deletion of the C-terminus or membrane-mediated mechanical stress, strongly affected the opening or closing behaviour of the loop D, and thereby the p_f .

4.2.4 Discussion

We have carried out molecular dynamics simulations to gain insights into the regulatory mechanism of the spinach plant aquaporin (SoPIP2;1). We quantified the water permeability of the channel in the two reported X-ray structural states, open and closed (figure 4.8). Our results revealed that in the open state, the channel conducts water at high rates, nearly two times the rate of hAQP1. This result is in excellent agreement with single-molecule

experiments (W. Kukulski and A. Engel, Biozentrum, private communication, 2008), and support therefore that the open X-ray structure would correspond to the physiological observed state of the protein in these assays. In addition, our simulation showed that in the closed state, the water permeability decreases one order of magnitude compared to the open state. So far, there are no water flux experiments reporting such a reduction in the water permeability. We predict, therefore, that if loop D undergoes a conformational change between the open and the closed X-ray states, as it is predicted from the current gating model, then a reduction in the p_f of at least one order of magnitude should be expected.

We observed that the permeability coefficient is correlated with the position of loop D along the principal vector connecting the closed and open X-ray states. Therefore, loop D would provide the necessary machinery for a proper gating mechanism. However, none of the investigated externally applied stimuli, i.e. phosphorylation of S115 and S274 (located in the cytosolic loop B and the C-terminus, respectively), protonation of H193 (located at the cytosolic loop D), deletion of the C-terminus, or membrane-mediated mechanical stress, strongly drove the loop D towards either the open or closed state, and thereby changed the water permeability, in a statistically significant manner. Our simulations, therefore, do not support a regulatory gating mechanism for the SoPIP₂;1 aquaporin mediated by phosphorylation of two serine residues or protonation of a histidine residue, as suggested in the current gating model, nor by membrane-mediated mechanical stress.

Equilibrium simulations, however, may suffer of insufficient sampling to capture the gating transitions, in particular, if this process occurs in a longer time scale than the simulated time. Consequently, future computational studies can address this issue by performing, for example, non-equilibrium driven molecular dynamics simulations. Accordingly, the impact of the mentioned external stimuli on the free energy associated with the gating process can be assessed in such computational studies. In addition, a hypothesis to be tested in future studies is whether phosphorylation plays a role in the trafficking of SoPIP₁;2, as has been already observed for other aquaporins such as AQP2 in renal cells [160, 161].

4.3 Is aquaporin 4 a voltage-regulated channel?

4.3.1 Introduction

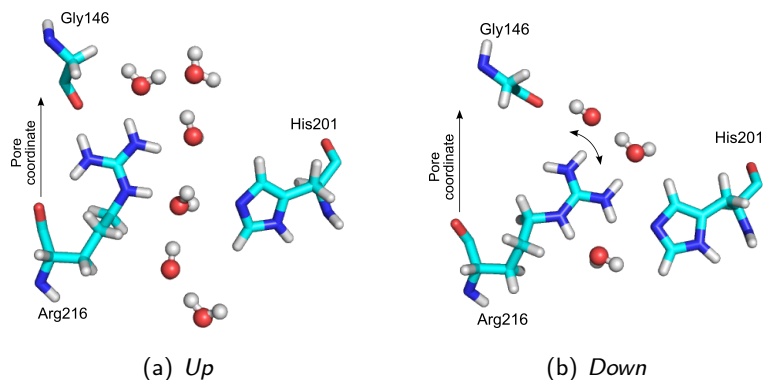


Figure 4.12. Two conformational states, up and down, adopted by the conserved arginine in the aromatic/arginine region. Figure adapted from the figure 5.4 of the simulation study carried out by Hub *et al.* for hAQP1 [87], changing the residue indices to the corresponding ones to hAQP4.

In a recent simulation study carried out by Hub *et al.* [87], a voltage-regulated gating mechanism was proposed for the human aquaporin 1 (hAQP1). In this study, the water permeability of hAQP1 was observed to be dependent on an applied electrostatic membrane potential.

The conserved arginine residue (Arg195 in hAQP1 and Arg216 in hAQP4) in the aromatic/arginine region (ar/R) emerged as the critical residue for the voltage regulation. In these simulations, this arginine was found to be flexible and frequently visited two conformational states, which differ in the dihedral angle along the C_{γ} - C_{δ} bond. The two states are visualized in figure 4.12 and in the following referred to as *up* and *down* state. In the *up* state, the arginine was stabilized by an intra-arginine hydrogen bond (H-bond) and by an H-bond to a glycine residue (Gly125 in hAQP1 and Gly146 in hAQP4) of Loop-C, allowing a continuous water file and rapid water flux. In the *down* state, the ar/R region was partly closed, the water file was interrupted, and a lever-like motion of the arginine (black arrow) side chain was required to allow the passage of a water molecule. These observations suggested that the openness of the ar/R region may be tuned mainly by direct electrostatic interactions of the arginine residue with the applied membrane potential.

As it was mentioned before, the arginine residue at the ar/R region is a highly conserved residue among the family of aquaporins. Moreover, the *up* state was found in the crystallographic structures of AQP1 and AQP4 [42, 43], but structural studies suggested that both, the *up* and the *down* state can be adopted in *E. coli* AQP-Z [53]. In addition, an alternative arginine position was found in hAQP1 from electron crystallographic studies [40], suggesting that

different arginine states, including the *up* and the *down* states, may indeed be populated under physiological conditions. All this evidence suggests therefore that other aquaporins may also be voltage regulated, by a similar mechanism as the one proposed by Hub *et al.* for hAQP1 [87]. Here we investigate this hypothesis taking as an example the human aquaporin 4 (hAQP4). We employed molecular dynamics simulations to study both the water permeation through hAQP4 and the arginine (Arg216 in hAQP4) conformational dynamics as a function of an applied electrostatic membrane potential.

4.3.2 Methods

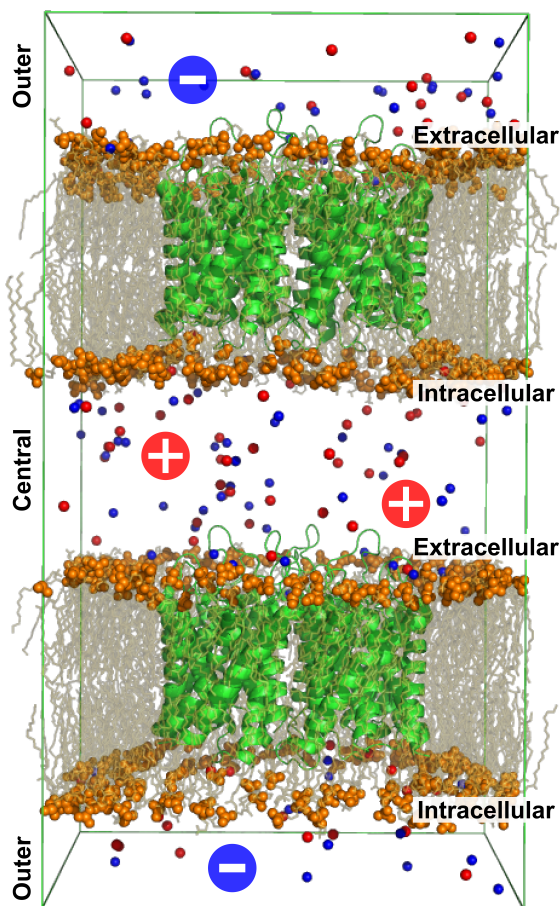


Figure 4.13. Double membrane simulation system to study the voltage regulation of hAQP4. Figure adapted from the figure 5.2 of Hub *et al.* [87]. The system is constituted by two POPE lipid bilayers (orange head groups and yellow tails), each one containing an hAQP4 tetramer (green), fully solvated by water (not shown). Due to the periodic boundary conditions, two water compartments, “central” and “outer”, are created between the the lipid bilayers. The two AQP tetramers are subject to an electrostatic membrane potential that is generated by placing an excess of cations (red) and anions (blue) in the central and outer compartments, respectively.

Simulation details

Molecular dynamics simulations were carried out by using a double membrane system as the one shown in figure 4.13, following the same approach explained in chapter 5 of Hub *et al.* [87]. The system is constituted by two fully solvated palmitoylcholine phosphatidylethanolamine

(POPE) lipid bilayers, each one containing an hAQP4 tetramer. Due to the periodic boundary conditions, two water compartments, “central” and “outer”, are created between the the lipid bilayers. The membrane potential is generated by placing an excess of cations and anions in the central and outer compartments, respectively. The two AQP tetramers are subject to a membrane potential of the same magnitude but different sign. Due to the limited size of the simulation system, the addition of one cation/anion pair induces a substantial $\Delta\Phi$ of ~ 200 mV. To allow for the simulation of smaller increments in $\Delta\Phi$, we placed, in addition to ions with integer charge, ions with fractional charges of $\pm e/2$ where e denotes the unit charge. Lennard-Jones parameters for the cations and anions with fractional charge were taken from the sodium and chloride ions, respectively.

The simulation boxes contain the two protein tetramers, 530 POPE lipids and 32420 TIP4P water molecules [140]. Crystallographic water molecules were kept in the structures, and approximately 150 mM sodium chloride was added to each water compartment. The AQP4 crystal structure was taken from the protein data bank (pdb code 3GD8 [162]). The OPLS all-atom force field [101, 102] was used for the protein, and lipid parameters were taken from Berger *et al.* [138]. The simulations were carried out by using the GROMACS simulation software [98, 99]. Electrostatic interactions were calculated with the particle-mesh Ewald method [114, 115]. Short-range repulsive and attractive interactions were described by a Lennard-Jones potential, which was cut off at 1.0 nm. The Settle algorithm [109] was used to constrain bond lengths and angles of water molecules, and Lincs [110] was used to constrain all other bond lengths. The fastest angular degrees of freedom involving hydrogen atoms were removed by using the virtual interaction-sites algorithm [111], allowing a time step of 4 fs. The temperature was kept constant by coupling the system to a velocity rescaling thermostat [116, 117] at 300 K with a coupling constant $\tau = 0.5$ ps. The pressure was kept constant by coupling the system to a semiisotropic Parrinello-Rahman barostat [122] at 1 bar with a coupling constant of $\tau = 5.0$ ps. The simulation length of the production runs was 70 ns discarding the first 5 ns as equilibration.

Water permeability calculations

The p_f was independently calculated for every monomer, based on the collective diffusion model [84] described in section 2.2.1. Water molecule displacements, dz_i , were computed every 10 ps, within a cylindrical region (of length $L = 0.84$ nm and radius $r = 0.5$ nm) between the NPA and the aromatic/arginine region, that is within the narrowest part of the channel. the mean square displacement of the collective variable n , $\langle n^2(t) \rangle$, was obtained by averaging over 650 time windows of 100 ps length each. Finally the p_f was obtained from the

slope of the curve $\langle n^2(t) \rangle$ versus time.

Calculation of the membrane potential

The electrostatic potential $\Phi(z)$, averaged over the simulation trajectory, was derived from the Poisson equation

$$\Delta\Phi(\mathbf{x}) = -\rho(\mathbf{x})/\epsilon_0 \quad , \quad (4.1)$$

where Δ denotes the Laplace operator, $\Phi(\mathbf{x})$ the electrostatic potential, $\rho(\mathbf{x})$ the time-averaged charge density, and ϵ_0 the vacuum permittivity. Accordingly, the charge density in the simulation was averaged over the $x - y$ plane and over the simulation frames, yielding a one-dimensional time-average density $\rho(z)$. $\Phi(z)$ was straight-forwardly computed by twice integrating $\rho(z)$. Because of the double integration, slightly non-converged charge densities may result in a non-physical offset in $\Phi(z)$ between the top and the bottom of the simulation box. In order to account for the periodic boundary conditions such offsets were removed by a linear correction in $\Phi(z)$.

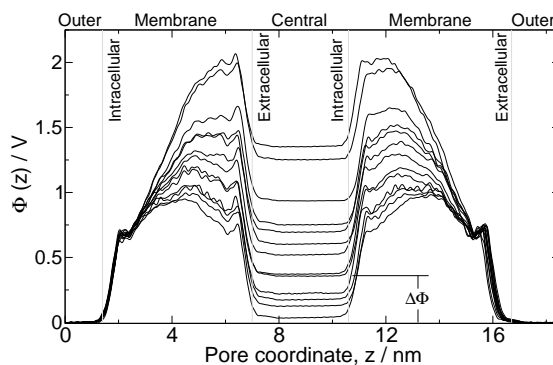
Distance between the arginine and the histidine in the ar/R region

The distance between the arginine and the histidine, d_{R-H} , in the aromatic/arginine (ar/R) constriction site was chosen to quantify the degree of closure of the channel. It was defined as the smallest distance between the C_ζ atom of the arginine to the nearest heavy atom of the histidine. d_{R-H} equals and 5.88 Å in the crystal structure of AQP4 [43].

Calculation of the splines and statistical errors

The smooth splines in figures 4.15 and the line in 4.16B were fitted to the data points from all individual monomers (not shown) using gnuplot. Statistical errors for the fitted smooth splines were computed using the Bayesian bootstrap [163]. Accordingly, the fit was carried out 200 times, with randomly selected weights for the individual data points for each fitting procedure. According to the Bayesian bootstrap, the weights are generated as follows [163]. Given n data points, draw $n - 1$ uniform random variables between 0 and n , and let $u_{(1)}, u_{(2)}, \dots, u_{(n-1)}$ denote their values in increasing order. In addition, let $u_{(0)} = 0$ and $u_{(n)} = n$. The random weights w_i are then defined by the gaps between two consecutive random numbers, i.e. $w_i = u_{(i)} - u_{(i-1)}$, where $i = 1, \dots, n$. The statistical error for the fit is given by the standard deviation computed from the 200 fitted splines. The statistical error in the linear least-square fit to P_{open} was computed using standard error propagation, based on the standard errors of

Figure 4.14. Electrostatic potential, $\phi(z)$, as a function of the pore coordinate, z , recovered from double membrane simulations of hAQP4. The membrane potential, $\Delta\phi$, is given by ϕ at the intracellular side minus ϕ at the extracellular side.



the slope and the y-intercept, which are derived by the fitting procedure. All statistical errors were plotted as 67% confidence intervals.

Statistical errors for the data points from individual simulations (black dots with dashed error bars in figures 4.15 and 4.16B) were computed from the the four values (p_f or P_{open}) taken from the four monomers via $\sigma/\sqrt{N_m - 1}$, where σ denotes the standard deviation and $N_m = 4$, the number of monomers. Because these errors are based on only four values, we stress they are not meant to represent a robust errors but rather as a semi-quantitative estimates, allowing to visualize the scattering of the monomeric values. Note that the statistical error in the fitted curves are substantially smaller than the error for the individual data points, because the fit effectively incorporates many data points at different membrane potentials, whereas the error for each data point is based on only $N_m = 4$ values.

4.3.3 Results and Discussion

Figure 4.14 presents the electrostatic potential $\Phi(z)$ as a function of the coordinate z (membrane normal), recovered from different double membrane simulations of hAQP4. The peaks in $\Phi(z)$ correspond to the intramembrane regions, and the flat parts in $\Phi(z)$ to the water layers. The membrane potential $\Delta\Phi$ is thus given by ϕ at the intracellular side minus ϕ at the extracellular side. The simulated membrane potentials lie in the range of -1.5 to $+1.5$ V, one order of magnitude larger than typical physiological potentials.

Figure 4.15 shows AQP4 single-channel water permeabilities p_f as a function of membrane potential. A p_f decrease was observed when switching from a positive to a negative membrane potential. Remarkably, a similar p_f -voltage dependence was observed by Hub *et al.* for hAQP1 [87]. Despite the substantial simulation time used to compute the p_f values, the individual values substantially scatter (error bars of the black dots). Hence, p_f converges relatively slowly with simulation time, suggesting that the large number of simulations employed here are indeed required to yield a robust p_f versus $\Delta\Phi$ signal. To guide the eye, we have fitted a

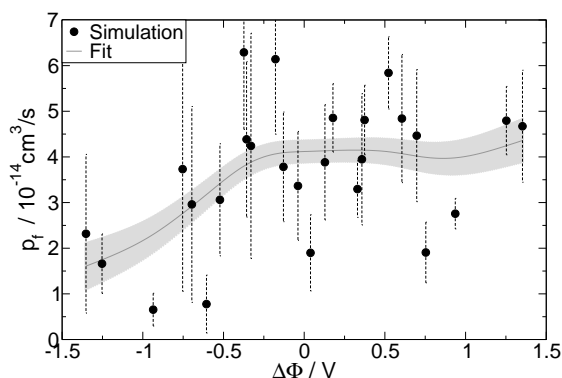


Figure 4.15. Water permeability coefficient as a function of an applied electrostatic membrane potential.

spline function to the data points (gray curve), where the shaded area indicates the statistical error of the fitted spline.

The voltage-sensitive openness of the aromatic/arginine (ar/R) region of hAQP4 was measured from the the distance d_{R-H} between Arg216 and His201 (figure 4.16). The distribution of d_{R-H} , taken from the 104 hAQP4 monomers in 13 simulations (figure 4.16a), resembles the sum of two Gaussians (dashed line). The conserved Arg216 adopts two distinct states, which agree with the *up* and *down* states observed in the hAQP1 simulations [87] (see figure 4.12). In the *up* state, the channel is open at the ar/R region. Accordingly, d_{R-H} in the crystal structure is at this region of the distribution (shaded bar). In contrast, in the *down* state, the conserved Arg216 occludes the pore and prevents water passage. The *up* state was predominantly visited when AQP4 was subject to positive membrane potentials (upper AQP tetramer in figure 4.13), and the *down* state to negative potentials (lower AQP tetramer in figure 4.13). To further quantify how voltage shifts the distribution towards either of the two states, the probability P_{open} for an open ar/R region ($d_{R-H} > 5.7\text{\AA}$) was computed (figure 4.16b). The probability correlates with $\Delta\Phi$, with the lowest P_{open} for negative $\Delta\Phi$, indicating a closed channel.

To estimate if the arginine displacement is sufficient to explain the tuning of P_{open} in response to $\Delta\Phi$, let us assume a simple open/closed two-state model. When switching between the *up* and the *down* state, Arg216 moved by $\delta z \sim 1.0\text{\AA}$ in the z-direction with respect to the center of mass of the respective monomer (data not shown). Let us in the following in addition assume a homogeneous electric field across the membrane and a membrane thickness of $d = 4\text{ nm}$. Then, $\Delta\Phi$ affects the potential energy difference between *up* and *down* state by $\Delta V(\Delta\Phi) = e \delta z \Delta\Phi / d$, where e denotes the unit charge. The relative populations of the open and closed states thus follow $P_{\text{open}}/P_{\text{closed}} = \exp(\Delta V/k_B T)$, where k_B and T denote the Boltzmann constant and the temperature, respectively. Using $P_{\text{open}} + P_{\text{closed}} = 1$, the

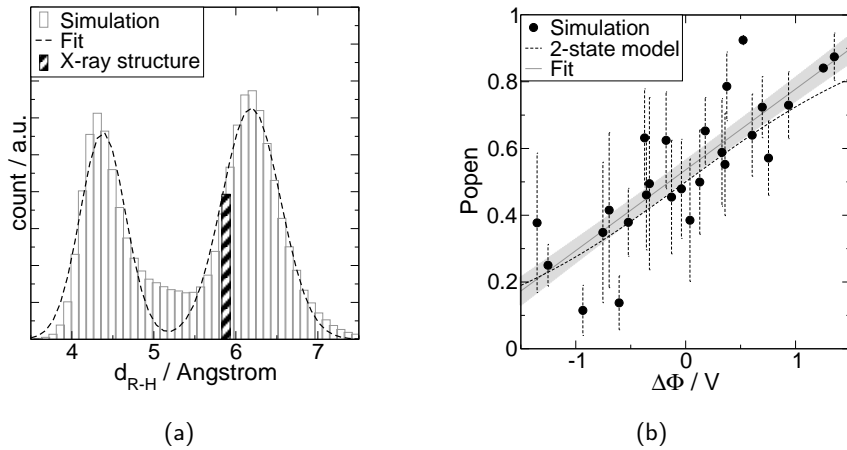


Figure 4.16. Voltage-sensitive openness of the aromatic/arginine (ar/R) region of hAQP4, as measured from the the distance d_{R-H} between Arg216 and His201. (a) Distribution of d_{R-H} , taken from 13 AQP4 simulations at membrane voltages between -1.5 and $+1.5$ V (gray histogram), revealing two distinct states. d_{R-H} in the AQP4 X-ray structure [43] is indicated by a shaded bar. (b) Probability for an open ar/R region ($d_{R-H} > 5.7\text{\AA}$) as a function of membrane potential $\Delta\Phi$ (black dots); a linear fit to the data points (gray lines), where shaded area indicates the statistical error, and P_{open} derived from a two-state model (dashed line).

probability of the open state is given by

$$P_{\text{open}}(\Delta\Phi) = \left[1 + \exp\left(-\frac{e\delta z}{dk_B T} \Delta\Phi\right) \right]^{-1} \quad (4.2)$$

$P_{\text{open}}(\Delta\Phi)$ is plotted as dashed black curves in figure 4.16(b). The curve is in reasonable agreement with the data points and the fitted lines, suggesting that electrostatic interactions of the arginine with the external field are indeed sufficient to explain the tuning of P_{open} .

4.3.4 Conclusion

We have carried out molecular dynamics simulations to study a putative voltage regulatory gating mechanism for hAQP4. We observed in our simulations that the single-molecule water permeability p_f of hAQP4 can be voltage regulated. In addition, the conserved arginine residue (arg216) at the aromatic/arginine region was found to be alternating between two states, gating the pore, and switching due to an applied electrostatic potential. These two findings are in remarkable agreement with the voltage-gated mechanism proposed by Hub *et al.* for hAQP1 [87], suggesting therefore that voltage sensitivity may be a general feature of AQPs. It will be highly interesting to test the simulation based voltage regulation hypothesis in AQPs experimentally.

Interaction of aquaporins with other (bio)molecules

Aquaporins do not work alone. They interact at different stages with different (bio)molecules such as proteins, lipids and organic compounds. During trafficking, aquaporins interact with other proteins (forming protein-protein complexes), to be efficiently translocated towards and from the cell membrane. After translocation, aquaporins are accommodated and stabilized at the lipid membrane by interacting with surrounding lipids, which also results in an alteration of the conformational properties of the lipid bilayer. Furthermore, aquaporins have been recognized as potential drug targets for the treatment of malaria and several water- and glycerol-imbalance related diseases, requiring a detailed understanding of the interactions of aquaporins with such a putative drugs (organic nontoxic compounds).

In this chapter, we focus on three important processes where aquaporins interact with different types of (bio)molecules. First, we study the formation and stability of the AQP2-LIP5 complex (aquaporin-protein interactions), which is a crucial process during the internalization of AQP2 from the membrane in renal cells (section 5.1). Second, we investigate the localization of DMPC lipids around AQP0, which is an excellent model to study protein-lipid interactions (section 5.2) . Third, we describe how molecular dynamics simulations can be used, in combination to experimental functional assays and molecular docking calculations, in the search and refinement of putative AQP9 blockers (section 5.3).

5.1 Aquaporin-protein interactions: predicting the structure of the complex formed by aquaporin 2 and the LIP5 protein.

5.1.1 Introduction

Water reabsorption takes place in the kidney as a body response to prevent dehydration [160]. During this process (reviewed in [160, 161]), water crosses across the cells which are separating the collecting duct system and blood vessels, to counteract changes in intravascular blood volume or in osmolality. The water crossing through the renal cells occurs in two steps: initially, triggered by the binding of the antidiuretic hormone vasopressin, aquaporin 2 (AQP2) proteins are translocated to the apical plasma membrane (facing the collecting duct), and water enters from the duct to the cells driven by an osmotic pressure. Subsequently, water exits the cells on the blood vessels side by permeating through AQP3 and AQP4. Once the water balance is recovered, the levels of the vasopressin hormone are reduced and AQP2 is internalized from the plasma membrane, and water entrance from the collecting duct is blocked. The trafficking of AQP2 to and from the plasma membrane is therefore a crucial step during the vasopressin-regulated renal water reabsorption.

Several proteins are implicated in the AQP2 trafficking. In particular, the lysosomal trafficking regulator protein-5 (LIP5) has been experimentally shown to interact with the C-terminal tail of AQP2, and suggested to facilitate the AQP2 internalization from the plasma membrane [88]. Further mutational studies demonstrated that the interaction between AQP2 and LIP5 is mainly mediated by three leucines in the C-terminus of AQP2: Leu230, Leu234 and Leu237 (P. Deen *et al.*, Nijmegen Medical Center, private communication, 2009).

The structure of LIP5 coupled to AQP2 (the AQP2-LIP5 complex) has not been determined yet. A key component of the structure of the complex is that the aquaporin tetramer is expected to be embedded in a vesicle membrane during trafficking. Moreover, equilibrium MD simulations of AQP2 (without LIP5), starting from an homology model based on the structure of AQP5, revealed that the C-terminal region, where Leu230, Leu234 and Leu237 are located, has a helical secondary structure, that is highly stabilized by direct contacts with the lipid environment (J. Hub, MPI for biophysical Chemistry, Göttingen, private communication, 2009). Thus, LIP5 is expected to bind to the AQP2 C-terminus in a such a way that is not overlapping with this part of the membrane and providing a binding site for this C-terminal helix.

Here, we use molecular dynamics simulations and related computational techniques to predict putative stable structures of the AQP2-LIP5 complex, mediated by interactions between the

C-terminal tail of AQP2 and LIP5, and compatible with the AQP2 tetramer embedded in a lipid bilayer.

5.1.2 Methods

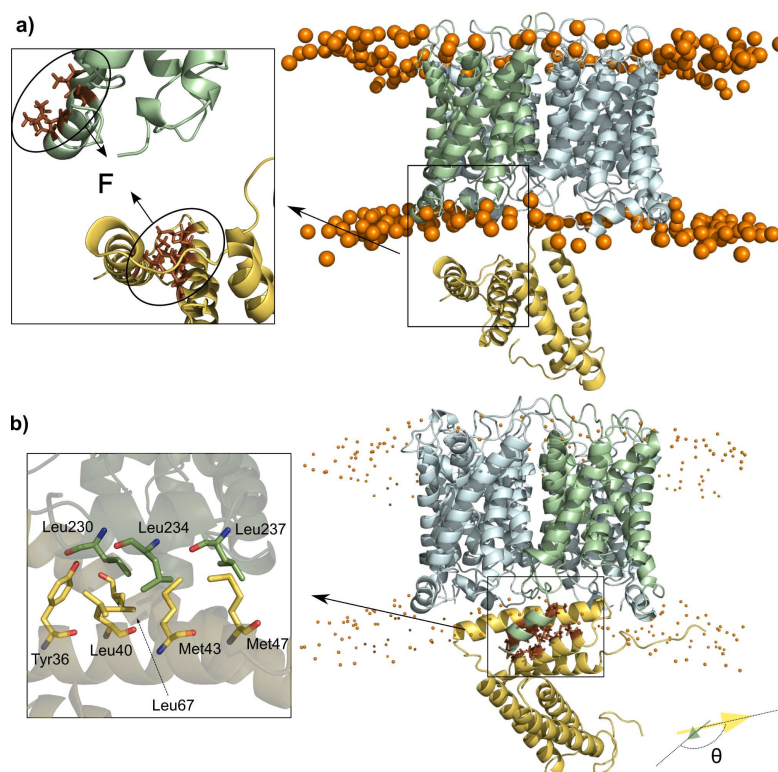


Figure 5.1. Molecular dynamics simulations of AQP2-LIP5 complex. The aquaporin tetramer (light cyan monomers A to C and green the monomer D) was embedded in a membrane of POPE lipids (phosphorus atoms shown in orange) and solvated by explicit water molecules (not shown). The LIP5 protein (yellow) was placed at the intracellular side. a) Generation of the complex by using pulling simulations. An additional harmonic force F (indicated with the arrows) was exerted on the three leucines at the C-terminal helix of the AQP2 D-monomer and the hydrophobic cleft of LIP5 (residues shown in brown). b) One of the three initial structures of the AQP2-LIP5 complex compatible with the membrane, generated by a combined Concoord-docking approach. The three leucines in the C-terminus of the AQP2 D-monomer (L230, L234, and L237) are located in a hydrophobic cleft of LIP5 (Y36, L40, M43, M47, and L67). The colored arrows on the right side indicate the orientation of the aquaporin C-terminal helix (green) with respect to the LIP5 hydrophobic cleft (yellow) measured by the angle θ .

Generating the initial conformation of the AQP2-LIP5 complex

Molecular dynamics simulations were carried out starting with the AQP2 tetramer in a fully solvated palmitoyl-oleoyl-phosphatidylethanolamine (POPE) lipid bilayer, and the LIP5 monomer intracellularly attached to one of the monomers of AQP2 (figure 5.1). Two independent ap-

proaches were used to generate initial structures of the AQP2-LIP5 complex. In the first approach, the LIP5 protein was pulled towards one AQP2 monomer in a 10 ns MD simulation, by adding a harmonic potential between the two proteins. In AQP2, the pulled residues were three leucine residues (Leu230, Leu234 and Leu237) located at the C-terminal helix of one of the AQP2 monomers (in the following referred to as AQP2 Ct-helix), whereas in LIP5, the pulled region was a hydrophobic cleft constituted by the residues Tyr36, Leu40, Met43, Met47, and Leu67, (in the following referred to as LIP5-cleft). The LIP5-cleft was chosen as a putative binding site based on a structural prediction of the complex using bioinformatic tools, carried out by H. Venselaar *et al.* (Nijmegen Medical Center, private communication, 2009). The pulling force was in the direction of the vector connecting the center of masses of the two groups of pulled atoms, and the harmonic force constant was $600 \text{ kJmol}^{-1}\text{nm}^{-2}$ (figure 5.1(a)).

In the second approach, a combination of conCOORD and docking calculations was used to generate the complex, and was carried out by D. Seeliger. Initially, an ensemble of conformations of an AQP2 monomer was generated, with its Ct-helix in different orientations with respect to the rest of the monomer, using the ConCOORD software [164]. Subsequently, the AQP2 monomer (in all the conCOORD-generated conformations) was docked to LIP5, by considering the Ct-helix as a "ligand" and the LIP5 protein as a "receptor". These docking calculations were carried out by using the Rosetta software [165, 166]. Finally, from the ensemble of docked conformations, with the AQP2 attached to LIP5 via the Ct-helix, three representative structures that did not show any substantial overlap with the membrane were selected (figure 5.1(b)).

MD simulations of the AQP2-LIP5 complex

Molecular dynamics simulations were carried out starting from the four putative structures of the AQP2-LIP5 complex: one from the structure generated by pulling simulations, after turning off the additional harmonic force between the Ct-helix in AQP2 and the hydrophobic cleft in LIP5, and the other three starting from the the structures generated by conCOORD+docking calculations.

The simulation boxes contain the aquaporin tetramer, the LIP5 monomer, 271 POPE lipids and around 27000 TIP4P water molecules [140]. The starting structure of the AQP2 and LIP5 were homology-modeled based on the x-ray structures of human AQP5 (PDB ID code 3D9S) [46] and of yeast Vta1 N-terminal domain (PDB ID code 2rkk) [167], respectively. The models were built by mutating differing residues by using bioinformatics tools, and carried out by H. Venselaar (Nijmegen Medical Center, private communication, 2009). Ions were also

added to neutralize the simulation systems.

The OPLS all-atom force field [101, 102] was used for the protein, and lipid parameters were taken from Berger *et al.* [138]. The simulations were carried out using the GROMACS simulation software [98, 99]. Long-range electrostatic interactions were calculated with the particle-mesh Ewald method [114, 115]. Short-range repulsive and attractive interactions were described by a Lennard-Jones potential, which was cut off at 1.0 nm. The Settle algorithm [109] was used to constrain bond lengths and angles of water molecules and Lincs [110] was used to constrain all other bond lengths. The fastest angular degrees of freedom involving hydrogen atoms were removed by using the virtual interaction-sites algorithm [111] allowing a time step of 4 fs, except in the 10 ns pulling simulation where the time step was 2 fs. The temperature was kept constant by coupling the system to a velocity rescaling thermostat [116, 117] at 300 K with a coupling constant $\tau = 0.5$ ps. The pressure was kept constant by coupling the system to a semiisotropic Parrinello-Rahman barostat [122] at 1 bar with a coupling constant of $\tau = 5.0$ ps. In the 10 ns pulling simulation, the Ct-helix of one of the aquaporin monomers was subject of additional restrains to preserve its secondary helical structure. The pulling simulation was previously equilibrated for 2 ns, with the coordinates of three CA residues of LIP5 harmonically restrained (with a harmonic force constant of $1000 \text{ kJmol}^{-1}/\text{nm}^2$), to maintain the initial distance between the AQP2 monomer and LIP5. Simulations starting from the concoord+docking structures were equilibrated for 500 ps before production. During this equilibration time the coordinates of the protein were harmonically restrained, with a harmonic force constant of $1000 \text{ kJmol}^{-1}/\text{nm}^2$. The simulation length of the production runs was between 210 ns and 250 ns.

Mutational studies

To study the role of the three Ct-helix leucines in AQP2 mediating the binding to LIP5, additional 200 ns simulations were carried out for the mutants that were known experimentally to prevent the complex formation: Leu230Val, Leu230Ile, Leu234Val-Leu234Val, and Leu234Val-Leu234Ile (P. Deen *et al.*, Nijmegen Medical Center, private communication, 2009). All these simulations were started from two independent structures of the AQP2-LIP5 complex: one obtained by the pulling simulations and the second one randomly selected from the three concoord-docking generated structures.

Observables

- **Intermolecular distance and orientation**

The distance between the centers of mass of the AQP2 Ct-helix and the hydrophobic

LIP5-cleft ($D_{\text{Ct-HC}}$) was monitored during the trajectory. In addition, the orientation of the AQP2 Ct-helix with respect to the LIP5-cleft was monitored, by measuring the angle θ formed by the vectors lying along the Ct-helix and the LIP5-cleft, respectively (figure 5.1(b)).

- **PCA analysis**

To monitor the orientation of the Ct-helix (with LIP5 attached to it) with respect to the aqp2 tetramer a principal component analysis was carried out [130] (see section 2.2.3). It was performed considering the coordinates of the backbone atoms of the C-terminal helix, accumulated over the whole trajectory, after least-squares fitting the remaining part of the monomer to the initial structure.

- **Rosetta docking-score calculations** A Rosetta [165, 166] docking score for the AQP2-LIP5 complex was computed for the wild type structures. This score is an empirical estimate of the binding free energy of the complex, and was carried out for 100 randomly selected snapshots (per trajectory) whose 2D pca projection was within the main visited pca region for that trajectory. These calculations were performed by D. Seeliger (MPI for Biophysical Chemistry, Göttingen, 2009).

5.1.3 Results and Discussion

The initial conformation of the AQP2-LIP5 complex

Four putative structures of the AQP2-LIP5 complex, with the aquaporin tetramer embedded in the membrane, were generated by two independent approaches: by pulling simulations or a concoord-docking based approach. In all of them, the Ct-helix of one aquaporin monomer, containing Leu230, Leu234, and Leu237, detaches from the membrane and was located on a hydrophobic cleft in LIP5 constituted by Tyr36, Leu40, Met43, Met43, and Leu67 (figure 5.1(b)). It is very unlikely that LIP5 binds to the aquaporin Ct-helix, when the latter is embedded in the membrane (as observed in equilibrium AQP2 MD simulations), because it would imply a substantial overlap of LIP5 with the membrane. Therefore, the detachment of the AQP2 Ct-helix from the lipid environment appears to be necessary for the binding to LIP5, and thereby for the formation of the complex. Interestingly, the docking calculations and bioinformatic analysis with the Yasara software (H. Venselaar *et al.*, Nijmegen Medical Center, private communication, 2009) independently predicted the same binding site in LIP5 (the LIP5-cleft), stressing therefore that this cleft is indeed the anchoring point in LIP5 for the complex formation. These structures therefore suggest that interactions between the three

C-terminal leucines of AQP2 and the hydrophobic cleft of LIP5 are stabilizing the complex, thus compensating for the unfavourable Ct-leucine solvent exposure created after the Ct-helix detaches from the membrane.

Dynamics of the AQP2-LIP5 complex

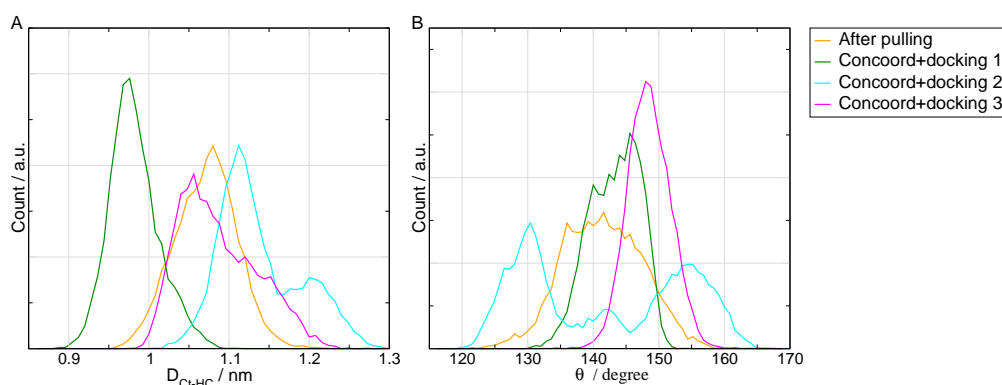


Figure 5.2. Intermolecular distance and orientation of the two anchor points in the AQP2-LIP5 complex, recovered from MD simulations starting from structures generated with two distinct methods: pulling simulations or a concoord+docking approach. (A) Histogram of the distance D_{Ct-HC} between the centers of mass of the three C-terminal leucines in AQP2 and the hydrophobic cleft in LIP5. (B) Histogram of the angle θ formed by the vectors lying along the aquaporin Ct-helix and the LIP5-cleft, respectively, as illustrated in figure 5.1(b).

The histogram of the distance D_{Ct-HC} between the centers of mass of the AQP2 Ct-helix and the hydrophobic LIP5-cleft, computed from the MD simulations, is shown in figure 5.2(A). In the four simulations, the AQP2 Ct-helix remained anchored to the LIP5-cleft during the whole simulated time (larger than 210 ns), with the side chains of the three leucines (Leu230, Leu234, and Leu37) located between the side chains of the hydrophobic residues in the LIP5-cleft (figure 5.1(b)). D_{Ct-HC} was found to fluctuate around an average value of 1.06 nm, with a standard error of 0.03 nm (taken from the four trajectory averages). Only in one of the simulations, starting from a concoord-docking based structure (cyan line), the distance showed an increment of approximately 1 Å in the last 70 ns, reflected in a second peak at around 1.2 nm in the distribution.

In addition, The angle θ formed by the vectors lying along the AQP2 Ct-helix and the LIP5-cleft was monitored, and the histograms are shown in figure 5.2(B). The average angle is 143° and the standard deviation σ for each simulation ranged between 3 and 12 degrees. In the simulation indicated by the cyan line, the separation between the Ct-helix and the LIP5-cleft (mentioned before) was accompanied by a rotation of approximately 25° , from 155° to 130° , thus leading to a bimodal distribution and consequently the largest standard deviation of 12° .

These results corroborate therefore the important role of Leu230, Leu234 and Leu237 located at the C-terminus in AQP2 for the formation and stability of the AQP2-LIP5 complex, as observed experimentally (P. Deen *et al.*, Nijmegen Medical Center, private communication, 2009). They also provide dynamical evidence for the hypothesis that the hydrophobic cleft in LIP5 (constituted by the residues Tyr36, Leu40, Met43, Met43, and Leu67) may serve as a binding region for the aquaporin Ct-helix, compensating for the solvent exposure of the three leucine residues, upon the Ct-helix detachment from the hydrophobic side of the membrane.

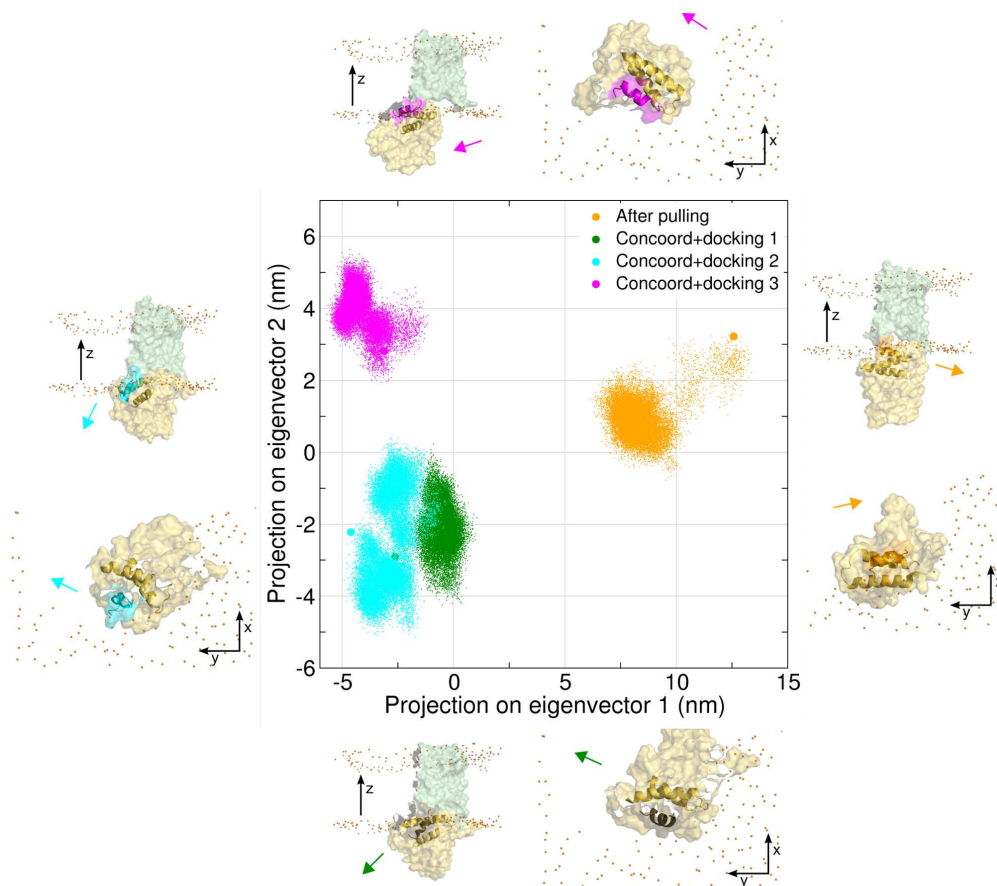


Figure 5.3. Projection onto the two principal eigenvectors of the wild type AQP2-LIP5 MD trajectories obtained by principal component analysis. Side and top views of a representative conformation of the AQP2-LIP5 complex are also displayed for the indicated simulations. The aquaporin monomer is shown in green and LIP5 in yellow. The C-terminal helix is colored as the indicated simulation. The colored arrows indicate the orientation of the C-terminal helix with respect to the aquaporin monomer, after aligning the monomer with the principal axes.

The orientation of the Ct-helix with respect to the rest of the AQP2 tetramer was also investigated by principal component analysis (figure 5.3). This orientation also gives possible positions of LIP5 with respect to the AQP2 tetramer, because the Ct-helix was observed

nearly rigidly attached to the LIP5-cleft. Two principal eigenvectors can be distinguished: eigenvector 1, showing to which direction the Ct-helix (and thereby the LIP5) is pointing in the plane parallel to the membrane (xy plane in figure 5.3), and eigenvector 2, that indicates the degree of tilting of the Ct-helix with respect to the z axis (normal to the membrane).

The simulation starting from the pulling-based structure (orange points) quickly moved away from the initial point (displayed with the orange circle) and later stabilized in another region of the 2D subspace. This indicates that the initial orientation of LIP5 with respect to AQP2 predicted in the pulling simulations was unstable, but also stresses that LIP5 could undergo rigid body motions tethered to the aquaporin Ct-helix.

In the simulation starting from the pulling-based structure, the Ct-helix (and thereby LIP5) points outwards the aquaporin tetramer (positive values of eigenvector 1). In contrast, in simulations starting from the concoord-docking based structures, the Ct-helix points inwards the aquaporin tetramer (negative values of eigenvector 1). In addition, two simulations starting from the concoord-docking based structure show a larger degree of tilting with respect to the z axis (negative values of eigenvector 2) compared to the simulation starting from the pulling based structure. In the latter, LIP5 thus accommodates more parallel to the membrane directly underneath the aquaporin tetramer.

Finally, there are three main clusters in the 2D pca subspace: cluster 1, shown in orange; cluster 2, corresponding to the regions colored in cyan and green, and cluster 3, depicted in magenta. This suggest that there are at least three putative conformations of the AQP2-LIP5 complex compatible with the AQP tetramer embedded in a membrane.

Effect of mutations of the C-terminal leucine residues

The effect of single and double mutations of the three mentioned AQP2 C-terminal leucines was also investigated. The average distance $\langle D_{\text{Ct-HC}} \rangle$ between the aquaporin Ct-helix and the LIP5-cleft was increased in the simulations with mutants, compared to the wild type simulations (figure 5.4(a)), although full detachment of LIP5 was not observed. In addition, the average angle θ between the two anchoring points was also severely affected by the mutants, with changes up to 20° for the double mutants L234V-L237I (grey points) and L234V-L237V (brown points) (figure 5.4(a)).

The orientation of the aquaporin Ct-helix (anchored to LIP5) with respect to the rest of the monomer was also modified by the mutations (see figure 5.4(b)). Remarkably, in the simulations starting from the pulling-based structure (upper-right clouds of points in each plot), mutations (coloured points) did not allow the trajectory to reach the final stable region reached in the wild type trajectory (black points). Moreover, in the simulation starting from

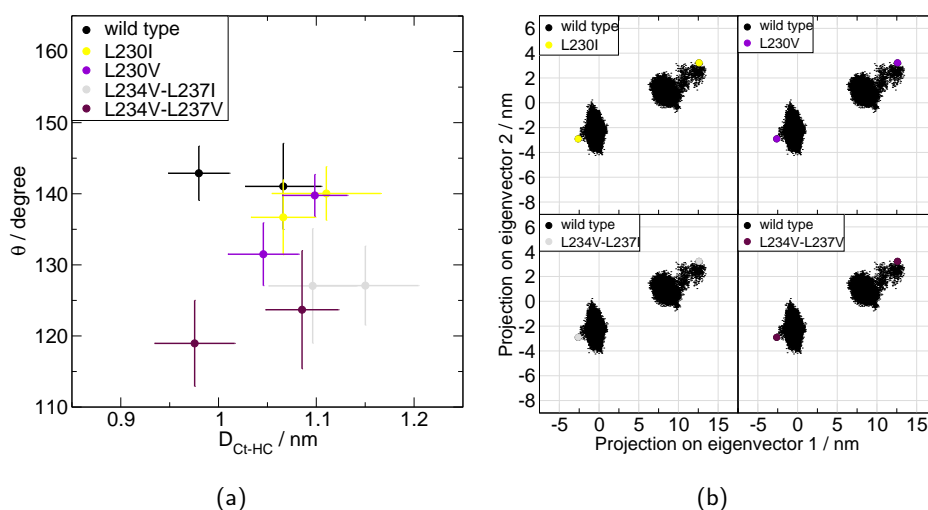


Figure 5.4. Effect of mutations of the C-terminal leucine residues. For each mutation two simulations were carried out, starting from the pulling-based or the concoord-docking structures. (a) Intermolecular distance and orientation of the aquaporin Ct-helix and the LIP5-cleft displayed as the average angle $\langle \theta \rangle$ as a function of the average distance $\langle D_{Ct-HC} \rangle$. The bars are the standard deviation indicating the width of both the distance and angle distributions. (b) Orientation of the Ct-helix (and thereby LIP5) with respect to the aquaporin tetramer quantified by the projection onto the two principal eigenvectors obtained by principal component analysis.

the concoord-based structure (lower-left clouds of points in each plot), the trajectories for the L234V-L237V mutant and the wild type protein converged to very different regions in the pca subspace.

In experimental assays, the studied mutations abolished the AQP2-LIP5 complex formation (P. Deen *et al.*, Nijmegen Medical Center, private communication, 2009). In our simulations with mutants, the LIP5 protein did not unbind from AQP2 as might be expected from the experiments. However, the mutants modified both the intermolecular distance and orientation between the anchoring points (Ct-helix in AQP2 and hydrophobic cleft in LIP5), as well as the position of LIP5 with respect to the aquaporin tetramer. Consequently, These events might correspond to the initial stages of the unbinding process. Our results therefore corroborate that mutations of Leu230, Leu234 and Leu237 at the C-terminus of AQP2 alter the binding affinity between AQP2 and LIP5, and suggest that the formation and abolishment of the complex occur on a much more slower time scale than the simulated time (200 ns).

Rosetta docking-score calculations

The average rosetta docking score and its standard deviation is presented in table 5.1. This value gives an empirical estimate of the binding free energy of the AQP2-LIP5 complex. Simulations labeled as 1 and 3 (orange and cyan in figure 5.3, respectively) gave the lower

Index	Simulation Initial structure generated by:	Rosetta score [a.u.] Average \pm st. dev.
1	pulling	-1067 \pm 9
2	concoord+docking	-1054 \pm 8
3	oncoord+docking	-1060 \pm 9
4	concoord+docking	-1051 \pm 8

Table 5.1. Average rosetta docking score computed extracting 100 representative snapshots from each each of the four wild type MD simulations.

score functions, suggesting that the ensemble of conformations in these two simulations have higher binding free energies than the ensembles in simulations 2 and 4, and therefore those structures of the complex would be energetically more favourable.

5.1.4 Conclusion and outlook

Here we have studied the formation, dynamics and stability of the AQP2-LIP5 complex by using molecular dynamics simulations. Our results predict at least two putative structures of the complex compatible with the aquaporin tetramer embedded in the membrane. In these structures, LIP5 adopts different orientations with respect to the aquaporin tetramer, tethered to the Ct-helix of one of the aquaporin monomers (figure 5.1(b)).

Our results corroborate the critical role of Leu230, Leu234 and Leu237, located at the aquaporin C-terminal helix, for the formation and stability of the AQP2-LIP5 complex. In addition, our data predict a cleft of hydrophobic residues (Tyr36, Leu40, Met43, Met43 and Leu67) in LIP5 as a potential binding site for the aquaporin Ct-helix. Our simulations therefore suggest a mechanism for the formation of the complex, in which the C-terminal helix in AQP2 detaches from the lipid environment and binds to this hydrophobic cleft in LIP5. Subsequently, attractive interactions between the three leucines located in this Ct-helix and the hydrophobic cleft in LIP5 would maintain the two proteins bound, compensating for the solvent exposure of the leucines, after they detach from the lipid environment.

To further distinguish which of the two putative predicted structures is more likely to occur, it would be highly interesting to experimentally determine the stoichiometry of the AQP2-LIP5 complex. This would allow to discard the structures showing clashes, when more than one LIP5 molecule binds to a single AQP2 tetramer. In addition, recent structural studies suggested that LIP5 dimerizes [167] and P. Deen *et al.* (Nijmegen Medical Center, private communication, 2010). However, whether LIP5 binds to AQP2 as a monomer or as a dimer remains to be elucidated. Recall that our simulations (with a single LIP5 monomer) would

be also compatible with a LIP5 dimer binding to AQP2, if the dimer adopts the structure obtained by Xiao *et al.* [167]. This is because, the LIP5 dimer still has the predicted binding site (the hydrophobic cleft) accessible for the binding to AQP2, and does not overlap with the membrane.

Furthermore, the impact of mutations at the hydrophobic cleft in LIP5 could be further investigated, both experimentally and in simulations, to confirm whether this region is indeed the binding site in LIP5. Finally, the energetics of the complex formation could be also addressed in further computational studies, to test if the free energy for the the aquaporin Ct-helix detachment from the lipid environment can be compensated by the subsequent binding to LIP5.

This study is expected to contribute in the understanding of the AQP2 trafficking, in particular during internalization steps, where LIP5 is suggested to play a role. In addition, this hopefully can be used as a starting model to further investigations on aquaporin-protein interactions.

5.2 Aquaporin 0: an excellent model to study protein-lipid interactions

5.2.1 Introduction

Biological lipid membranes are the natural environment of aquaporins and other membrane proteins. Understanding protein-lipid interactions is therefore crucial to elucidate how membrane proteins are stabilized within the lipidic environment and also how they alter the conformational properties of lipids.

AQP0 is the most abundant membrane protein in lens fiber cells. There, it acts as an anchor of thin junctions [168], in addition to its function as a water channel [169, 170]. AQP0 has emerged as an excellent model to study protein-lipid interactions. This is mainly motivated by recent electron crystallography studies of AQP0 ([38, 39] and reviewed in [171]), that revealed the structure of AQP0 with some annular lipids surrounding it: first, with dimyristoylphosphatidylcholine (DMPC) lipids (at a resolution of 1.9 Å) [38], and second, with *E. Coli* polar (EPL) lipids [39] (at a resolution of 2.5 Å).

These structures gave structural insights on how AQP0 accommodates into different lipid bilayers and how annular lipids adapt to the AQP surface. They also raised the following questions: first, how representative are the observed lipid structures of those adopted by lipids when aquaporin tetramers are not as densely packed as in the 2D crystals used for the structure determination?. Second, which lipidic groups (heads or tails) do interact with AQP0?. Third, how do annular lipids mediate between the protein and the pure bulk lipids?. Here, we addressed these questions by calculating time-average lipid density maps around the AQP0 protein, from molecular dynamics simulations of an AQP0 tetramer embedded in membrane of DMPC lipids.

5.2.2 Methods

Molecular dynamics simulations were carried out starting with the AQP0 tetramer in a fully solvated Dimyristoylphosphatidylcholine (DMPC) lipid bilayer (Fig. 5.5). Two independent simulations were carried out: Including the crystallographic lipids observed in the electron microscopy (EM) structure of AQP0 [38], and excluding such crystallographic lipids.

The simulation boxes contain the protein tetramer, 278 and 288 DMPC lipids in the simulation with and without the crystallographic lipids, respectively, and around 23400 TIP4P water molecules [140]. The AQP0 structure was taken from the Protein Data Bank (PDB ID code 2B6P [38]). The tetramer was inserted into the lipid bilayer by using the `g_membed` software [137]. Crystallographic water molecules were kept in the structures and ions were added to

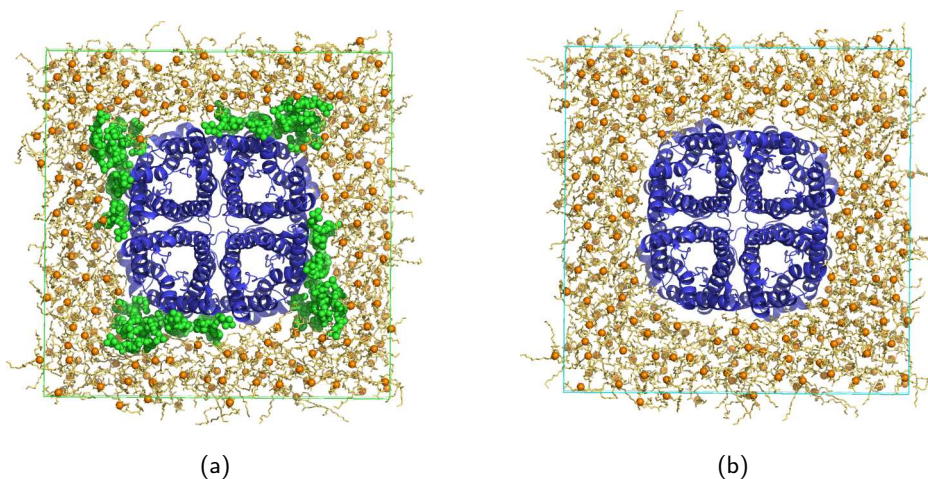


Figure 5.5. Molecular dynamics simulations of AQP0 embedded in a patch of DMPC lipids. Top views of the simulation box illustrating the tetramer (blue), embedded in a DMPC lipid bilayer (orange) and solvated by water (not shown). Two simulations were carried out: (a) Including the crystallographic lipids (shown in green) observed in the electron microscopy structure of AQP0 [38], and (b) excluding such crystallographic lipids.

neutralize the simulation systems. The OPLS-AA [101, 102] all-atom force field was used for the protein, and lipid parameters were taken from Berger *et al.*[138]. The simulations were carried out using the GROMACS simulation software [98, 99]. Long-range electrostatic interactions were calculated with the particle-mesh Ewald method [114, 115]. Short-range repulsive and attractive interactions were described by a Lennard-Jones potential, which was cut off at 1.0 nm. The Settle algorithm [109] was used to constrain bond lengths and angles of water molecules and Lincs [110] was used to constrain all other bond lengths. The fastest angular degrees of freedom involving hydrogen atoms were removed by using the virtual interaction-sites algorithm[111], allowing a time step of 4 fs. The temperature was kept constant by coupling the system to a velocity rescaling thermostat [116, 117] at 300 K with a coupling constant $\tau = 0.1$ ps. The pressure was kept constant by coupling the system to a semiisotropic Berendsen barostat [116] at 1 bar with a coupling constant of $\tau = 1.0$ ps. All simulations were equilibrated for 4 ns before production. During this equilibration time the coordinates of the protein were harmonically restrained, with a harmonic force constant of $1000 \text{ kJmol}^{-1}/\text{nm}^2$. The simulation length of the production runs was 100 ns.

Lipid density maps around the AQP0 tetramer

Lipid density maps were calculated to study the arrangement of the DMPC lipids around AQP0. The density maps obtained from electron microscopy represent the electrostatic potential of the atoms, also known as the shielded Coulomb potential [172]. This quantity

(defined in the real space) is correlated with the atomic scattering factors (defined in the reciprocal space):

$$f(k = \sin \theta / \lambda) = \sum_{i=1}^4 a_i e^{-(b_i+B)k^2/4} + c_1 e^{-Bk^2/4}. \quad (5.1)$$

Here, $f(k)$ is fitted to four gaussian functions, with a_i , b_i and c_1 fitting parameters depending on the atom type, and the B -factor is explicitly taken into account. The atomic potential $\rho(r)$ was calculated by the Fourier transformation of $f(k)$ [172]:

$$\rho(r) = \sum_{i=1}^4 \frac{a_i \sqrt{\pi}}{\sqrt{b_i + B}} e^{-\frac{4\pi^2 r^2}{b_i + B}} + \frac{c_1 \sqrt{\pi}}{\sqrt{B}} e^{-\frac{4\pi^2 r^2}{B}}. \quad (5.2)$$

Due to the four fold symmetry of the aquaporins, each monomer has identical lipid interface, therefore, $\langle \rho \rangle$ was considered only around a single AQP0 monomer. The trajectories of the four monomers (with their surrounding lipids) were concatenated, after fitting the monomer to the reference crystallographic monomeric position. $\langle \rho \rangle$ was constructed by time-averaging over the resulting 4×100 ns trajectory. $\rho(t)$ was calculated at every time step t , in a 3-dimensional grid of dimensions $7 \times 7 \times 6$ nm³, with a resolution of 0.4 \AA , and centered in the center of mass of the monomer. The density at the i -th point of the grid was estimated adding the ρ contributions (given by equation 5.2) of the atoms nearby the grid point:

$$\rho_i(t) = \sum_{j \in \text{cut off}} \rho(|\mathbf{R}_i(t) - \mathbf{r}_j(t)|), \quad (5.3)$$

where $\mathbf{R}_i(t)$ and $\mathbf{r}_j(t)$ are the coordinates of the i -th grid point and the j -th atom, respectively. Here, the sum was carried out only with the atoms within a cut off distance of 0.3 \AA to the i -th grid point. a_i , b_i and c_1 were taken from Hirai *et al.* [172]. The atomic potentials remained practically unchanged for two different choices of the B -factor: 1 \AA^2 and 20 \AA^2 . Thus a value of $B = 20 \text{ \AA}^2$ was considered. The maps were drawn and analyzed with the PyMOL software [123].

5.2.3 Results and discussion

Lipid arrangement around AQP0

The average lipid density maps $\langle \rho \rangle$ around an AQP0 monomer computed from MD simulations is presented in figure 5.6. Remarkably, both simulations including or excluding the crystallographic lipids (observed in the EM structure of AQP0) show similar density maps. This suggests that lipids quickly adopted stable positions around AQP0, in a time scale of 100 ns, and independently on the initial conditions: either taking the lipid positions from the

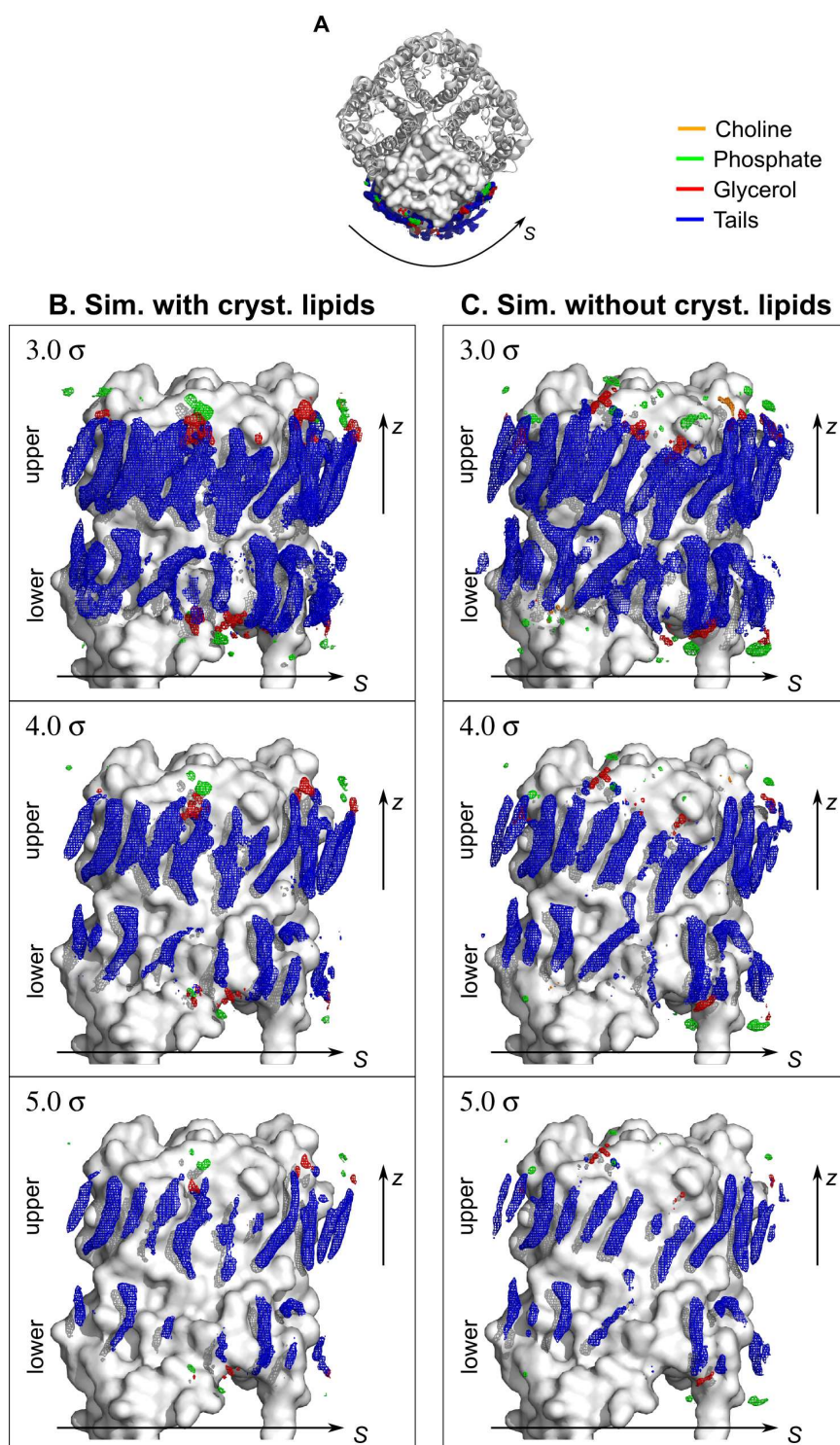


Figure 5.6. Average lipid density maps $\langle \rho \rangle$ around an AQP0 monomer computed from MD simulations. Due to the four fold symmetry of the aquaporin tetramer, each monomer has identical lipid interface, allowing for the calculation of $\langle \rho \rangle$ around a single monomer. The contribution of each part of the lipids is shown with different colors (see labels). A. Top view of the AQP0 tetramer and $\langle \rho \rangle$ around a single monomer. B. and C. are side views of $\langle \rho \rangle$ for the simulation with and without considering the crystallographic lipids, respectively. Here, z is the coordinate normal to the membrane (indicating the lower and the upper leaflet) and L is the coordinate surrounding the AQP0 monomer indicated in A. Each panel presents the map contoured at three different σ values.

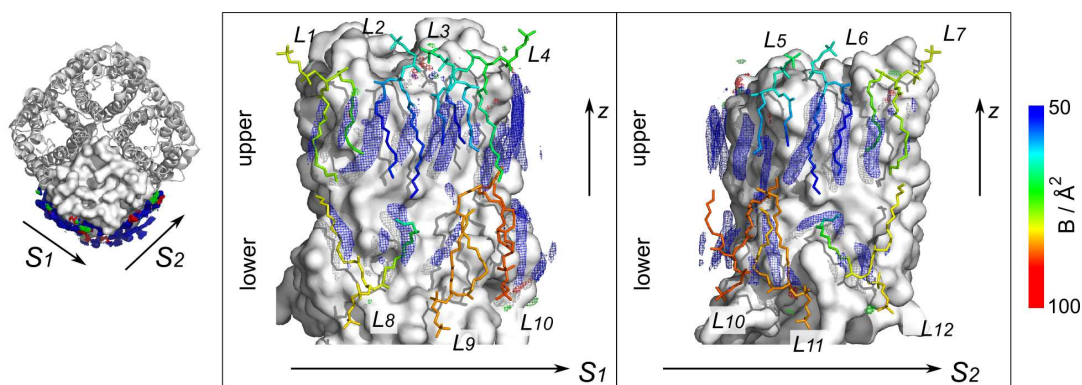


Figure 5.7. Comparison between the calculated lipid density maps (ρ) and the lipid structures determined by electron crystallography [38]. Maps were derived from the simulation without including such crystallographic lipids (Fig. 5.5(b)). Maps are countoured at 5.0σ (blue mesh) and crystallographic lipids are depicted in stick representation (according to the B-factor color scale at the right side). The panels in squares correspond to the side views S_1 and S_2 indicated in the top view of the tetramer at the left side.

EM crystallographic lipid structures or randomly from the equilibrated patch where AQP0 was inserted. High density regions (with well defined contours up to 5.0σ) were observed mainly for the lipid tails (blue mesh). These strong densities indicate positions around AQP0 where the lipids tails were predominantly found, accommodating the AQP0 surface roughness. In contrast, more diffuse densities were found for the head groups (choline, phosphate and glycerol), suggesting that the head groups of the annular lipids have a high mobility or static disorder, and therefore can be discarded as specific lipid-AQP0 interaction sites. Interestingly, the region between the two leaflets had also weak densities, indicating that the end of the tails also have a high degree of mobility.

The density map showed an asymmetry between the two leaflets, concerning the number of well defined tail positions. In the upper leaflet the map revealed 10 positions, that were mostly occupied by 5 lipids, or with a lower population by 6 lipids with two lipids sitting perpendicular to the AQP0 surface each one contributing with one tail. In the lower leaflet only 6 tail positions were observed, with much weaker densities than the ones in the upper leaflet. This asymmetry may be attributed to the shape of the AQP0 monomer, narrower in the lower leaflet than in the upper leaflet, leaving more space for the lower lipids to move.

Comparison with the crystallographic lipid positions

A comparison between the density maps calculated from MD simulations and the lipid structures derived by electron crystallography [38] is presented in figure 5.7. The computed density maps are in reasonable agreement with the crystallographic lipid positions. Most of the crys-

tallographic lipids (stick representation) had some portions of their tails matching with the high density regions of the calculated map (blue mesh). In particular, tails of lipid 6 in the upper leaflet almost fitted entirely within the strong density places; the middle part of lipid 1 and the tail end of lipid 12 also corresponded to high density regions, and the separation of the tails of lipid 8 were also captured by the map. Our simulations therefore support that the positions of lipid 1, 6 and 8 are representative of those formed by lipids around single AQP0 tetramers embedded in a DMPC lipid bilayer.

Remarkably, the measured B-factors showed a favourable degree of correlation with the computed densities. The portion of the lipids with low B-factor values (less uncertainty in the positions) matched with the high density regions. Accordingly, the high B-factor lipid parts (more uncertainty in the positions) are located in regions of weak or diffuse densities.

Crystallographic lipids were observed packed between AQP0 tetramers in 2D crystals of AQP0 [38]. Thus, lipids facing the side S_1 of certain tetramer are also facing the side S_2 of its adjacent tetramer. In consequence, crystallographic lipids in side S_1 are specular reflections of the ones in side S_2 , from the surface dividing S_1 and S_2 . In our simulations, carried out with a single tetramer, this is not the case. This may therefore explain why the high density countours (associated with stable lipid positions around AQP0) and the crystallographic lipid positions did not match at both sides S_1 and S_2 at the same time: e.g. lipid 1 (in side S_1) matched while the lipid 7 (in side S_2) did not, and the same for the lipids 6 and 8 compared to lipid 2 and 12, respectively.

Bulk lipid behavior recovered when moving away from AQP0

The densities for lipids distant from the AQP0 monomer was also monitored in our simulations (figure 5.8). When moving away from the AQP0 protein, the density gradually smeared out, and lipids started to behave as bulk lipids. The strong density regions are only observed for annular lipids. Interestingly, at intermediate distances (few layers away of the AQP0 monomer), a moderate localized behaviour of the lipids is still observed (in particular for the upper leaflet shown in the first two panels from the top to the bottom). This suggests that the high localization of the first layer of annular lipids influence the positioning of the neighbor lipid layers. This effect gradually vanishes as the lipids start to separate of the AQP0 protein, until the bulk lipid behaviour is finally recovered.

5.2.4 Summary and outlook

Here, we have studied the localization of DMPC lipids around AQP0, by calculating average lipid density maps from molecular dynamics simulations. Our simulations revealed that the

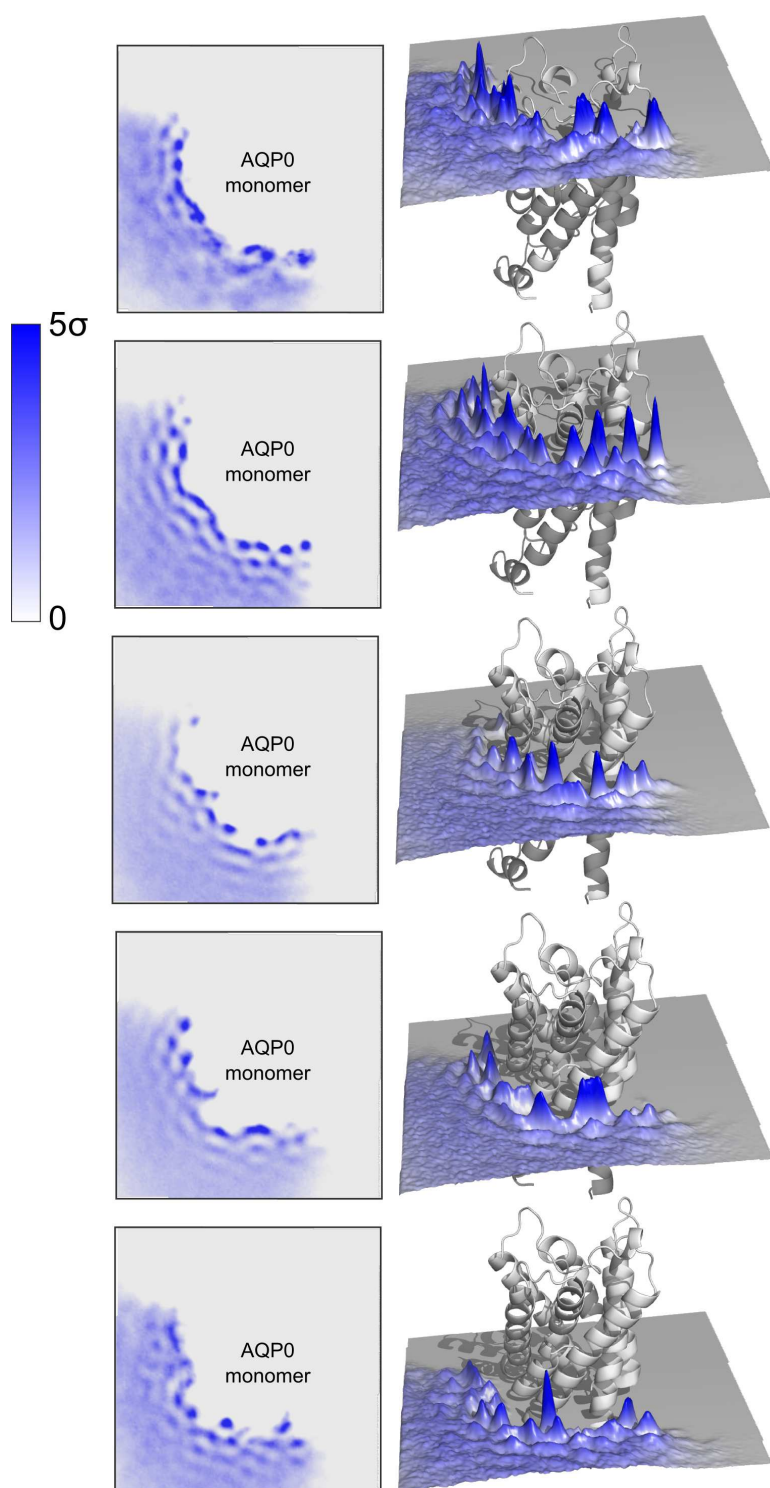


Figure 5.8. Lipid density maps at different heights of the AQP0 monomer, derived from the simulation excluding the crystallographic lipids. The left panels represent the density (in color gradient representation) at the different transversal cuts indicated in the right panels. In the right panels, both the color and height in the map represent the lipid density.

tails of the annular lipids presented the highest localized positions, accommodating to the roughness of the AQP0 surface. In contrast, the head lipid groups had weaker densities, stressing on their high degree of mobility or disorder and suggesting that they play only a minor role as stabilizing lipid-protein interaction sites. Our results therefore support the hypothesis that AQP0-lipid interactions are mainly mediated by the matching of the lipid tails to the AQP0 surface, as suggested from electron microscopy structural studies [38, 39]. The calculated maps are in reasonable agreement with the crystallographic electron microscopy lipid positions, supporting that the the crystallographic positions of at least three lipids (1, 6 and 8 in figure 5.7) are representative of those adopted by annular lipids in contact with AQP0 tetramers, at low AQP0 concentrations where tetramers are not densely packed. In addition, the favourable agreement also provides an independent validation of the used lipid and protein simulation parameters (force field).

In our simulations, lipids were found to gradually recover their bulk-lipid properties when moving away from AQP0. Furthermore, the lipids facing the AQP0 surface influenced the localization of the neighbour surrounding lipid layers. Consequently, lipids in contact with AQP0 would provide an interface between the pure bulk lipids and the protein as pointed out by Gonen et al. [38].

This preliminary study is expected to motivate future research in the understanding of protein-lipid interactions. The effect of the lipid tail length and lipid composition on the protein-lipid interactions could be further investigated, by following a similar computational approach, and compare with the recent structural studies of AQP0 with *E. Coli* polar lipids [39]. In addition, the effect of cholesterol could be also incorporated in the simulations. Furthermore, the density maps derived here could be used for refinement of the current crystallographic lipid positions observed in electron microscopy.

5.3 Inhibition of aquaporin 9

5.3.1 Introduction

Malfunction of aquaporins is associated with several diseases (reviewed in [5, 63–65]). On the one hand, the distorted expression of aquaporins was found to be implicated in diseases such as nephrogenic diabetes insipidus in the kidney [66], inherited cataracts in the eye [67, 68], or deficient secretion of saliva associated with the Sjögren's syndrome [69]. On the other hand, excessive water transport through aquaporins was also described as a pathological condition in diseases such as congestive heart failure and cirrhosis [70, 71]; brain [72, 73] and lung [74] edema; glaucoma [63, 64, 75], and lung cancer [76].

In addition to water permeation, aquaporins were found to be involved in disorders in glycerol metabolism, such as obesity [77]. Furthermore, the *Plasmodium falciparum* malaria aquaglyceroporin (PfAQP) has been suggested to play crucial roles during the parasitic life cycle [25, 79] (as pointed out in chapter 3).

Aquaporins are therefore promising drug targets for the treatment of malaria and other diseases associated with water- or glycerol-imbalance. In particular, specific AQP9 blockers are potential candidates to control disorders in glycerol metabolism [77], and may prevent the propagation of malaria in the initial phases of infection [89, 90].

Here, we describe how molecular dynamics simulations can be used, in combination to experimental functional assays and molecular docking calculations, in the search and refinement of putative AQP9 blockers. We identified an initial set of compounds that both showed an inhibitory effect on AQP9 (measured experimentally) and had a high virtual docking score when binding to AQP9. Subsequently, the dynamical and functional properties of the compounds were analyzed in molecular dynamics simulations, in terms of their structural stability at binding sites in the AQP9 pore predicted by docking calculations and their effect on AQP9 water permeability.

5.3.2 Methods

Selection of putative blocker compounds

Blocker candidates were selected based on two criteria: their inhibitory effect measured in assays containing AQP9, and their computed docking score when binding to the AQP9 protein. Initially, 124 organic ligands were selected from a database of chemical compounds. Their inhibitory effect was assessed by measuring the shrinking time of cells containing AQP9 in the presence of such compounds. These experiments were carried out by M. Rützler (University of Aarhus, Aarhus, Denmark). An increase in the shrinking time was assumed to be associated

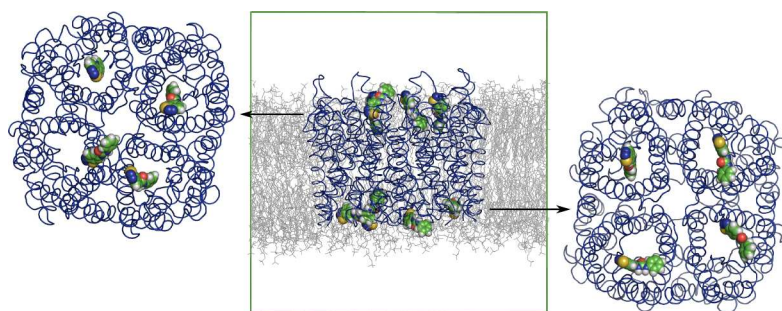


Figure 5.9. Molecular dynamics simulations of AQP9 with putative blocker compounds. The middle panel displays a side view of the simulation box containing the aquaporin tetramer (blue) embedded in a patch of DPPC lipids (grey) and solvated by water (not shown). In addition, putative blocker molecules (green spheres) were included, bound to each monomer either intra- or extra-cellularly. The initial binding pose of these molecules was generated by molecular docking calculations. Compounds were docked either at the extracellular or the intracellular vestibules (right and the left figures depicting top and bottom views, respectively).

to the inhibition power of the compound. Thus, the compounds were sorted from large (high inhibition) to small (low inhibition) shrinking times.

These 124 ligands were also docked to two different AQP9 conformations, extracted from an MD simulation of AQP9 (see below for details of the modeling of the initial AQP9 structure). These docking calculations yielded the putative ligand binding sites in the AQP9 protein and their docking scores, and were carried by S. Wäcker (MPI for Biophysical Chemistry, Göttingen) using the FlexX software [173]. The ligands were docked to each one of the four AQP9 monomers (each one having a different conformation), either at the extra- or the intracellular vestibule, giving eight putative binding poses (per compound) with their corresponding docking scores (see Fig. 5.9). Compounds were ranked according to their docking scores in four different ways: first, according to the best of the eight score values; second, according to the average extracellular score value (when the compound was bound at the extracellular vestibule); third, the same as the second item but when the compound was bound to the intracellular vestibule, and fourth, taking the best between these two "extracellular" and "intracellular" average score values. By averaging the individual monomeric docking scores, corresponding to different AQP9 monomer conformations, the effect of the conformational flexibility of the AQP9 protein was taken into account.

Finally, those compounds that both had an inhibitory effect (observed in the functional assays) and were ranked among the 20 best places (by any of the four ranking ways) in the two different AQP9 conformations were considered for further molecular dynamics simulations studies.

Molecular dynamics simulations

Molecular dynamics simulations were carried out starting from an AQP9 tetramer embedded in fully solvated Dipalmitoylphosphatidylcholine (DPPC) lipid bilayer, and the blocker candidates bound to the AQP9 tetramer (Fig. 5.9). Twelve different compounds were considered for these MD simulations. Two independent simulations were carried out for each compound, bound to the aquaporin tetramer either extra- or intra-cellularly. In addition, a simulation of the AQP9 protein without bound compound was performed for comparison.

The simulation boxes contain the protein tetramer, 4 ligand molecules, 271 DPPC lipids, and around 23470 TIP3P water molecules [140]. The starting structure of AQP9 was modeled based on the x-ray structure of the *E. Coli* glycerol facilitator (GlpF) (PDB ID code 1FX8), [50]), by mutating differing residues by using the WHAT IF modeling software [136]. The tetramer was inserted into the lipid bilayer by using the g_membed software [137]. Ions were added to neutralize the simulation systems. Initial compound positions, bound either at the extracellular or the intracellular vestibule of each aquaporin monomer, were derived from the docking calculations mentioned above. The amber99SB [103] all-atom force field was used for the protein, and lipid parameters were taken from Berger *et al.*[138]. Parameters for the putative blockers were derived from the generalized amber force field (GAFF) by using the antechamber packages [174, 175]. Parametrization of these compounds was carried out by S. Wäcker (MPI for Biophysical Chemistry, Göttingen).

The simulations were carried out using the GROMACS simulation software [98, 99]. Long-range electrostatic interactions were calculated with the particle-mesh Ewald method [114, 115]. Short-range repulsive and attractive interactions were described by a Lennard-Jones potential, which was cut off at 1.0 nm. The Settle algorithm [109] was used to constrain bond lengths and angles of water molecules and Lincs [110] was used to constrain all other bond lengths, allowing a time step of 2 fs. The temperature was kept constant by coupling the system to a velocity rescaling thermostat [116, 117] at 300 K with a coupling constant $\tau = 0.1$ ps. The pressure was kept constant by coupling the system to a semiisotropic Parrinello-Rahman barostat [122] at 1 bar with a coupling constant of $\tau = 5.0$ ps. All simulations were equilibrated for 1 ns before production. During this equilibration time the coordinates of the protein were harmonically restrained, with a harmonic force constant of $1000 \text{ kJmol}^{-1}/\text{nm}^2$. The simulation length of the production runs was 100 ns.

Water permeability calculations

The single-channel osmotic permeability, p_f , was independently calculated for every monomer, based on the collective diffusion model [84] described in section 2.2.1. Water molecule dis-

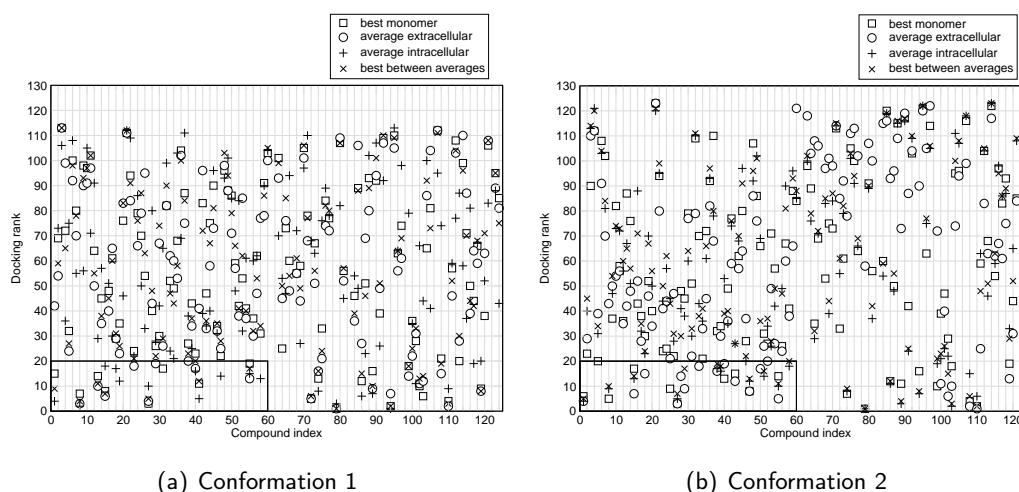


Figure 5.10. Docking ranking for a set of putative AQP9 inhibitors, derived from molecular docking calculations. (a-b) Docking rank calculated for two different AQP9 structures (extracted from a control MD simulation), by using four different methods (indicated in the plot labels and explained in the text). Compounds are sorted along the x axis according to the measured shrinking time, from large (high inhibition) to small (low inhibition) values. The first 60 compounds were observed to increase the shrinking time compared to the time measured in control assays (in the absence of compounds). Those compounds that both showed an increment in the shrinking time and were ranked between the 20 best positions (by any of the four ranking methods) for the two AQP9 conformations were selected for further MD simulations. They are indicated in the plots by the square region at the lower left side.

placements, dz_i , were computed every 2 ps, within a cylindrical region (of length $L = 2.05$ nm and radius $r = 0.5$ nm) centered at the pore axis and spanning -0.7 nm down and 1.35 nm up from the NPA region. $\langle n^2(t) \rangle$ was obtained by averaging over 350 time windows of 200 ps length each. Finally the p_f was obtained from the slope of the curve $\langle n^2(t) \rangle$ versus time. An effective p_f value was obtained by averaging the values of the four monomers and the error was estimated as the standard error.

5.3.3 Preliminary results

The docking ranking for a set of putative AQP9 inhibitors derived from molecular docking calculations is presented in figure 5.10. Compounds are sorted in the x axis according to the measured shrinking time, from large (high inhibition) to small (low inhibition) values. As can be seen, the docking ranking shows little correlation with the measured inhibitory effect. This could be attributed to two facts: first, in the functional assays, the compounds may not be blocking the AQP9 water channels, but altering the shrinking time by a different mechanism instead, or, second, in the docking calculations, the docking score is a very rough estimate of the binding free energy, and only two conformations of the AQP9 protein may not be

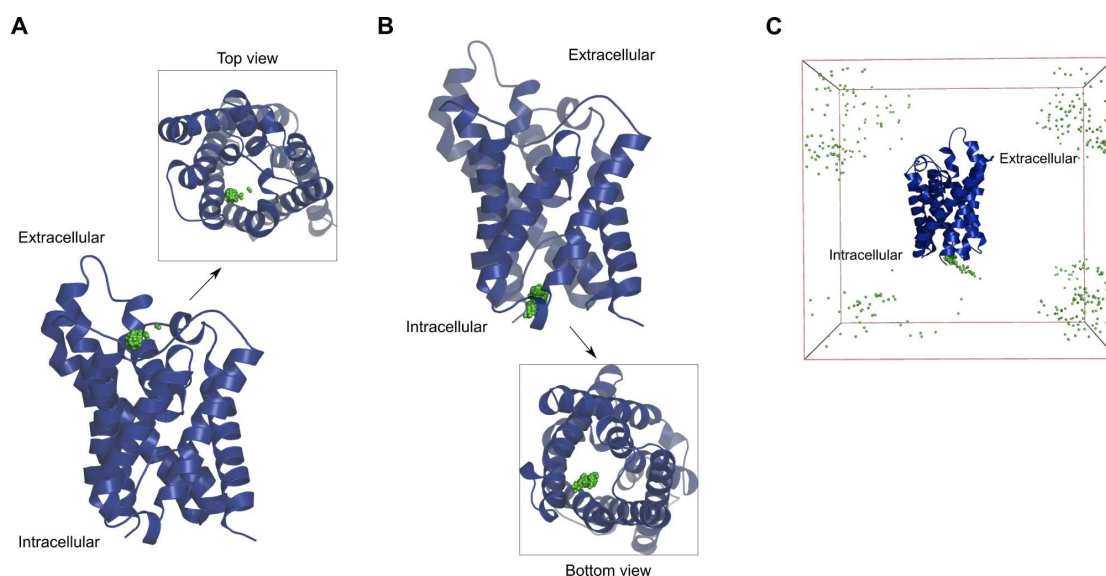


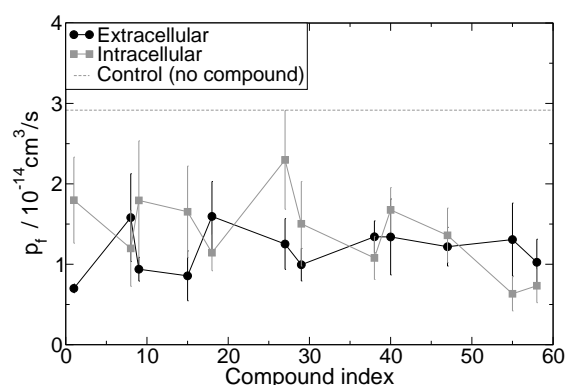
Figure 5.11. Dynamics of the compounds described by molecular dynamics simulations, remaining bound to the AQP9 monomer extracellularly (A) or intracellularly (B), or unbinding the AQP9 protein and diffusing in the water media (C). The AQP9 monomer is depicted in blue and the positions of the center of mass of the compound visited during the simulation are indicated by green dots. The insets in A. and B. show top and bottom views of the simulations with the compound bound extracellularly or intracellularly, respectively. In C. the compound visits different corners of the simulation box due to the periodic boundary conditions.

sufficient for an accurate description of the energetics of binding. Nevertheless, these data can still be used as a preliminary filtering criterium, to select putative compounds for further calculations, such as molecular dynamics simulations. In fact, twelve compounds that both showed an inhibitory effect (seen in an increase in the shrinking time) and had the highest ranking docking positions were selected for further MD simulation analysis.

Molecular dynamics simulations of the compounds bound to AQP9 allowed to identify stable binding sites both at the extracellular and the intracellular vestibules, as depicted for one of the compounds in figures 5.11A and 5.11B. In contrast, some of the compounds quickly unbound the AQP9 monomer and diffused away in the aqueous medium (figure 5.11C), indicating a low binding affinity at those places of the AQP9 monomer (unbinding events already taking place in a time scale of 100 ns), and suggesting that such binding sites could be discarded for further analyzes and as putative blockers.

Figure 5.12 shows AQP9 single-channel permeabilities p_f in the presence of the putative blocking compounds. All compounds reduced the water permeability of AQP9, compared to the value computed in a control simulation in the absence of compounds (dashed horizontal grey line). For each compound, the p_f values differ depending whether the compound binds either at the extracellular side (black dots) or the intracellular side (grey squares). Therefore,

Figure 5.12. Water permeability coefficients of AQP9 in the presence of putative blocking compounds. Compounds 27 and 55 unbound of one of the four monomers at the intracellular side. However a reduction in the average p_f was still observed in these two cases, due to the fact that in the other three monomers the compound remained bound.



the p_f calculations provide a method to classify the binding of the compounds at particular regions of the protein, according to their effect on the water permeation. Compounds 1, 9 and 15, for example, severely reduced the AQP9 water permeability, binding at the extracellular vestibule, whereas they only induced a moderate reduction binding at the intracellular side. Compounds 27 and 55 unbound of one of the four monomers at the intracellular side. However a reduction in the average p_f was still observed in these two cases, due to the fact that in the other three monomers the compound remained bound.

5.3.4 Summary and outlook

Here, we have illustrated how molecular dynamics simulations can be used, in combination to experimental functional assays and molecular docking calculations, in the search of putative AQP9 blockers.

Experimental functional assays and molecular docking calculations provided a set of twelve putative AQP9 blockers, and predicted their binding sites inside the AQP9 protein. Molecular dynamics simulations of these compounds bound to the AQP9 protein corroborated the stability of some of the predicted binding places. In addition, these simulations allowed to discard those predicted sites with low binding affinity, where the compound quickly unbound the protein in a time scale of 100 ns.

The studied compounds induced a reduction in the single-channel AQP9 water permeability, compared to a control simulation in the absence of compounds. These calculations also indicated whether the compounds, in addition to being stably bound at certain positions inside the protein, also altered the water permeability. Thus, these calculations provided a method to quantify how the binding of certain compounds affects the water permeability of a single AQP9 molecule. This information can not be directly assessed in functional assays, where the measured shrinking time also depends on the AQP9 copy number, and the compounds may be altering the membrane permeability by a different indirect mechanism rather than blocking

the AQP9 channels.

This study can be continued with further analysis of the MD simulation trajectories to identify residues of AQP9 which are specifically interacting with the ligand, and that could be tested in (both experimental and computational) mutational studies to validate the predicted binding mode. Moreover, the chemical similarity between the studied compounds can also be addressed, in order to distinguish the essential chemical groups (of the compounds) interacting with the AQP9 protein. Knowledge of these chemical groups is expected to inspire the design and refinement of novel compounds that specifically bind to AQP9 and efficiently inhibit its solute permeation.

Finally, a similar approach, combining experiments and simulations, can be followed in order to search specific compounds binding to other disease-implicated aquaporins.

Summary and conclusions

Aquaporins are protein channels responsible for the permeation of water and other small solutes through biological membranes. They play a vital role for cells by maintaining water homeostasis, allowing the uptake or expulsion of water, in response to external osmotic pressure. Aquaporins are extremely efficient molecular nanodevices, with several thousands of millions ($\sim 10^9$) of water molecules diffusing through the channel per second, while maintaining strict selectivity, only permitting the passage of small uncharged solutes.

Aquaporins have been subject of intense research over the last two decades. A large number of experimental and computational studies have contributed to explain how aquaporins perform their function, how they are implicated in crucial physiological processes in different living organisms, and how several diseases are associated with their malfunction. In particular, molecular dynamics (MD) simulations have played a protagonist role, providing dynamic and energetic information on aquaporins that is usually difficult to access experimentally, such as single-channel water permeabilities or the free energy for the permeation of solutes.

In the present thesis, we have used MD simulations to expand our understanding on the molecular machinery of aquaporins. We addressed the following goals by employing MD and related computational methods: first, to provide a quantitative description of the solute permeation through the *Plasmodium falciparum* aquaglyceroporin, second, to investigate the molecular determinants governing aquaporin gating, and third, the study of interactions of aquaporins with other (bio)molecules, such as proteins, lipids and organic compounds. Our findings, conclusions and perspectives for future work are summarized in the following sections.

Solute permeation through the *Plasmodium falciparum* aquaglyceroporin

The aquaglyceroporin from *Plasmodium falciparum* (PfAQP) is a potential drug target for the treatment of malaria. It efficiently conducts water and other small solutes, and is proposed to participate in several crucial physiological processes during the parasitic life cycle. The first goal of the present thesis was to provide a dynamic and energetic description at the single-molecule level of the solute permeation through PfAQP, by using MD simulations. Initially, the water permeation through PfAQP was quantified by computing the single molecule osmotic permeability coefficient. Subsequently, the pore geometry and the energetics for water transport through PfAQP were analyzed by computing the radius and potential of mean force profiles from equilibrium simulations. Finally the energetics of permeation of glycerol and urea was studied by computing potentials of mean force by using the technique of umbrella sampling simulations.

Our simulations confirm that PfAQP is a highly efficient water channel, that is able to conduct water at single-molecule permeability rates comparable to the rates for the human aquaporin-1 (hAQP1) and the *Escherichia coli* glycerol facilitator (GlpF). Furthermore, we identified the hydrophobic regions near the NPA motif as the main water rate limiting barriers. The Arg196 residue at the aromatic/arginine (ar/R) region, the most constricted part of the pore, was found to play a crucial role regulating the permeation of water, glycerol and urea. The computed free energy barriers at the ar/R selectivity filter corroborate that PfAQP conducts glycerol at higher rates than urea, and suggest that PfAQP is a more efficient glycerol and urea channel than GlpF.

Our results are consistent with a solute permeation mechanism for PfAQP which is similar to the one established for other members of the aquaglyceroporin family. In this mechanism, hydrophobic regions near the NPA motifs are the main water rate limiting barriers. In addition, the replacement of water-arg196 interactions and solute-matching in the hydrophobic pocket at the ar/R region are the main determinants underlying selectivity for the permeation of solutes like glycerol and urea.

This study is anticipated to guide further computational and experimental studies in the search of putative blockers of PfAQP. The understanding of their mechanism of action can hopefully be used to interrupt crucial physiological processes of the malaria parasite such as water regulation and glycerol uptake.

The gating mechanism of aquaporins

Regulation of aquaporins has nowadays become an active field of research. Channel gating has emerged as an efficient regulatory mechanism, facilitating a rapid response to external

stimuli when other regulatory mechanisms, such as transcriptional regulation or trafficking, are too slow [54]. Accordingly, the second goal addressed in this thesis was to make use of MD simulations to investigate the molecular determinants of gating of the yeast aquaporin (Aqy1), the spinach plant aquaporin (SoPIP1;2), and the human aquaporin 4 (hAQP4).

- **Gating of yeast aquaporin mediated by phosphorylation or mechanosensing.**

Our results suggest that the yeast aquaporin-1 (Aqy1) may be regulated by both phosphorylation of a serine residue (Ser107) or mechanosensing. Both possibilities lead to similar opening transitions after a local rearrangement of the residues Tyr31, Leu189, Ala190 and Val191, located at the gate of the pore. We observed that there is a principal collective motion causally involved in these gating transitions, and that it is possible to attain reproducible opening events along this collective coordinate. Our results also suggested, for the mechanism of mechanosensing, that external forces triggering gating transitions are transmitted from the lipid membrane to the gate of the pore (which is not in direct contact with the membrane) via coupled movements of the helices four, five, and six, the latter being in direct contact with the membrane.

Our simulations together with structural and functional studies support a mechanism in which both phosphorylation and mechanosensitive gating can trigger the channel opening. Regulation of Aqy1 by phosphorylation may also be exploited in physiological contexts, when less rapid changes in water transport activity are required. Hence, phosphorylation may fine tune the water flux during normal conditions of growth, whereas mechanosensitive gating could provide a rapid pressure valve in response to unexpected shocks. In addition, Aqy1 regulation may help yeast to survive rapid freezing and thawing, aiding the organism's quest to adapt and survive.

- **Is Spinach plant aquaporin gated by phosphorylation or histidine-protonation?**

We carried out MD simulations to gain insights into the regulatory gating mechanism of the spinach plant aquaporin (SoPIP2;1). We studied how the water permeability and the conformation of the cytosolic loop D (the proposed gate of the pore) are affected by the two stimuli proposed in the current gating mechanism: first, phosphorylation of two serine residues (located in the cytosolic loop B and the C-terminus, respectively), and second, protonation of a histidine residue (in the loop D). We also analyze the impact of the deletion of the C-terminus on these two functional and structural observables. Finally, we explore the possibility that SoPIP2;1 might be gated by membrane-mediated mechanical stress.

We quantified the water permeability of the channel in the two states, open and closed,

which differ in the structure of loop D, as observed by X-ray crystallography. Our results revealed that in the open state, the channel conducts water at high rates, nearly two times the rate of human aquaporin-1, which is in excellent agreement with single-molecule experiments, supporting therefore that the open X-ray structure would correspond to the physiological state of the protein observed in these assays. In addition, our simulations showed that in the closed state, the water permeability decreases one order of magnitude compared to the open state. So far, there are no water flux experiments reporting such a reduction in the water permeability. We predict, therefore, that if loop D undergoes a conformational change between the open and the closed X-ray states, as it is proposed from the current gating model, then a reduction in the p_f of at least one order of magnitude should be expected.

We observed that the permeability coefficient is correlated with the position of loop D along the principal vector connecting the closed and open X-ray states. Therefore, loop D would provide the necessary machinery for a proper gating mechanism. However, none of the investigated externally applied stimuli, i.e. phosphorylation of S115 and S274 (located in the cytosolic loop B and the C-terminus, respectively), protonation of H193 (located at the cytosolic loop D), deletion of the C-terminus, or membrane-mediated mechanical stress, strongly drove the loop D towards either the open or closed state, and thereby changed the water permeability, in a statistically significant manner. Our simulations, therefore, do not support a regulatory gating mechanism for the SoPIP2;1 aquaporin mediated by phosphorylation of two serine residues or protonation of a histine residue, as suggested in the current gating model, nor by membrane-mediated mechanical stress.

Equilibrium simulations, however, may suffer of insufficient sampling to capture the gating transitions, in particular, if this process occurs in a longer time scale than the simulated time. Consequently, future computational studies can address this issue by performing, for example, non-equilibrium driven molecular dynamics simulations. Accordingly, the impact of the mentioned external stimuli on the free energy associated with the gating process can be assessed in such computational studies. In addition, a hypothesis to be tested in future studies is whether phosphorylation plays a role in the trafficking of SoPIP1;2, as has been already observed for other aquaporins such as AQP2 in renal cells [160, 161].

- Voltage-regulation of human aquaporin-4.

We explored the hypothesis of voltage-regulation for human aquaporin-4 (hAQP4) *in silico*. We observed in our simulations that the single-molecule water permeability p_f of hAQP4 can be voltage regulated, with a p_f decrease when switching from a positive to a negative

membrane potential. The conserved arginine residue (arg216) at the aromatic/arginine region was found to be alternating between two states, gating the pore, and switching due to an applied electrostatic potential. These two findings are in remarkable agreement with the voltage-gated mechanism proposed by Hub *et al.* for hAQP1 [87], suggesting therefore that voltage sensitivity may be a general feature of AQPs. Voltage-regulation of hAQP4 may have important physiological role in the central nervous system (CNS). There, hAQP4 colocalizes with the Kir4.1 potassium channel, and both proteins have been suggested to work as a water-potassium transport complex [73]. It will therefore be highly interesting to test the simulation based voltage regulation hypothesis in AQPs experimentally and to study its physiological implications at the CNS.

Interaction of aquaporins with other (bio)molecules

Aquaporins do not work alone. They interact at different stages with different (bio)molecules such as proteins, lipids and organic compounds. As a third focus of the present thesis, we studied three important processes where aquaporins interact with different types of (bio)molecules. First, we studied the formation and stability of the AQP2-LIP5 complex (aquaporin-protein interactions), which is a crucial process during the internalization of AQP2 from the membrane in renal cells. Second, we investigated the localization of DMPC lipids around AQP0, which is an excellent model to study protein-lipid interactions. Third, we described how MD simulations can be used, in combination with experimental functional assays and molecular docking calculations, in the search and refinement of putative AQP9 blockers.

- The complex formed by aquaporin-2 and the LIP5 protein.

In renal cells, the translocation of aquaporin-2 (AQP2) is a process of vital importance. The lysosomal trafficking regulator protein-5 (LIP5) has been experimentally shown to interact with the C-terminal tail of AQP2, and suggested to facilitate the AQP2 internalization from the plasma membrane [88]. By employing MD simulations, we studied the formation, dynamics and stability of the AQP2-LIP5 complex. Our results predict at least two putative structures of the complex compatible with the aquaporin tetramer embedded in the membrane. In these structures, LIP5 adopts different orientations with respect to the aquaporin tetramer, tethered to the Ct-helix of one of the aquaporin monomers.

Our results corroborate the critical role of Leu230, Leu234 and Leu237, located at the aquaporin C-terminal helix, for the formation and stability of the AQP2-LIP5 complex. In addition, our data predict a cleft of hydrophobic residues (Tyr36, Leu40, Met43, Met43 and Leu67) in LIP5 as a potential binding site for the aquaporin Ct-helix. Our simulations

therefore suggest a mechanism for the formation of the complex, in which the C-terminal helix in AQP2 detaches from the lipid environment and binds to this hydrophobic cleft in LIP5. Subsequently, attractive interactions between the three leucines located in this Ct-helix and the hydrophobic cleft in LIP5 would maintain the two proteins bound, compensating for the solvent exposure of the leucines, after they detach from the lipid environment.

In future studies, to further distinguish which of the two putative predicted structures is more likely to occur, it will be highly interesting to experimentally determine the stoichiometry of the AQP2-LIP5 complex. This will allow to discard the predicted structure where the LIP5 is more tilted with respect to the AQP2 monomer, not allowing to bind additional LIP5 molecules to the other AQP monomers. Furthermore, the impact of mutations at the hydrophobic cleft in LIP5 can be investigated, both experimentally and in simulations, to confirm whether this region is indeed the binding site in LIP5. Finally, the energetics of the complex formation can be also addressed in future computational studies, to test if the free energy for the the aquaporin Ct-helix detachment from the lipid environment can be compensated by the subsequent binding to LIP5.

- **Aquaporin-0: an excellent model to study protein-lipid interactions.**

Biological lipid membranes are the natural environment of aquaporins and other membrane proteins. Understanding protein-lipid interactions is therefore crucial to elucidate how membrane proteins are stabilized within the lipidic environment and also how they alter the conformational properties of lipids. AQP0 has emerged as an excellent model to study protein-lipid interactions, mainly motivated by recent electron crystallography studies [38, 39] that revealed the structure of some annular lipids around the AQP0 protein. We have studied the localization of DMPC lipids around AQP0, by calculating average lipid density maps from MD simulations.

Our simulations revealed that the tails of the annular lipids presented the highest localized positions, accommodating to the roughness of the AQP0 surface. In contrast, the lipid head groups were found to display weaker densities. Our results therefore support the hypothesis that AQP0-lipid interactions are mainly mediated by the matching of the lipid tails to the AQP0 surface, as suggested from electron microscopy structural studies [38, 39].

The calculated maps are in reasonable agreement with the crystallographic electron microscopy lipid positions. This supports that some of the crystallographic positions are indeed representative of those adopted by annular lipids around AQP0 tetramers immersed at low concentrations in a lipid bilayer. In addition, the favourable agreement also provides an independent validation of the used lipid and protein simulation parameters (force field).

In our simulations, lipids were found to gradually adopt bulk-lipid properties when located further away from AQP0. Furthermore, the lipids facing the AQP0 surface influenced the localization of the neighbour surrounding lipid layers. Consequently, lipids in contact with AQP0 appear to provide an interface between the pure bulk lipids and the protein as pointed out by Gonen et al. [38].

In future research, the effect of the lipid tail length and lipid composition on the protein-lipid interactions can be further investigated, by following a similar computational approach, and comparing with the recent structural studies of AQP0 with *E. Coli* polar lipids [39]. In addition, the effect of cholesterol can be also incorporated in the simulations. Furthermore, the density maps derived here may be used for refinement of the current crystallographic lipid positions observed in electron microscopy.

- **Inhibition of aquaporin-9.**

Aquaporins are promising drug targets for the treatment of malaria and other diseases associated with water- or glycerol-imbalance. In particular, specific AQP9 blockers are potential candidates to control disorders in glycerol metabolism [77], and may prevent the propagation of malaria in the initial phases of infection [89, 90]. Here, we described how MD simulations can be used, in combination with experimental functional assays and molecular docking calculations, in the search and refinement of putative AQP9 blockers.

Experimental functional assays and molecular docking calculations provided a set of twelve putative AQP9 blockers, and predicted their binding sites inside the AQP9 protein. MD simulations of the these compounds bound to the AQP9 protein corroborated the stability of some of the predicted binding sites, and allowed to discard those with low binding affinity for further analysis.

The studied compounds induced a reduction in the single-channel AQP9 water permeability, compared to a control simulation in the absence of ligands. These calculations also indicated whether the compounds, in addition to being stably bound at certain positions inside the protein, also altered the water permeability. Thus, these calculations provided a method to quantify how the binding of certain compounds affects the water permeability of a single AQP9 molecule. This information can not be directly assessed in functional assays, where the measured shrinking time also depends on the AQP9 copy number, and the compounds may be altering the membrane permeability by a different indirect mechanism rather than blocking the AQP9 channels.

This study can be continued with further analysis of the MD simulation trajectories to identify residues of AQP9 which are specifically interacting with the ligand. The importance

of these residues for ligand binding can be tested in (both experimental and computational) mutational studies to validate the predicted binding mode. Moreover, the chemical similarity between the studied compounds can also be addressed, in order to distinguish the essential chemical groups interacting with the AQP9 protein. This knowledge is expected to inspire the design and refinement of novel compounds that specifically bind to AQP9 and efficiently inhibit its solute permeation.

This thesis creates a basis for future research on the permeation phenomena through biological membranes mediated by aquaporin channels. In addition, it can hopefully be used in the search and design of drugs for the treatment of malaria and other diseases associated with the aquaporin malfunction, and in the development of elaborated nanodevices with technological applications.

Bibliography

- [1] G. M. Preston, T. P. Carroll, W. B. Guggino, and P. Agre. Appearance of water channels in xenopus oocytes expressing red cell chip28 protein. *Science*, 256:385–387, 1992.
- [2] B Albert, A Johnson, J Lewis, M. Raff, K Roberts, and P. Walter. *Molecular Biology of the Cell*. Garland Science, 2002.
- [3] Alan Finkelstein. *Water movement through lipid bilayers, pores, and plasma membranes*. John Wiley & Sons, 1987.
- [4] Bert L de Groot and Helmut Grubmüller. Water Permeation Across Biological Membranes: Mechanism and Dynamics of Aquaporin-1 and GlpF. *Science*, 294(5550):2353–2357, 2001.
- [5] Landon S. King, David Kozono, and Peter Agre. From structure to disease: the evolving tale of aquaporin biology. *Nat Rev Mol Cell Biol*, 5(9):687–698, September 2004.
- [6] J. Bernard Heymann and Andreas Engel. Structural clues in the sequences of the aquaporins. *Journal of Molecular Biology*, 295(4):1039 – 1053, 2000.
- [7] Rafael Zardoya. Phylogeny and evolution of the major intrinsic protein family. *Biol. Cell*, 97(6):397–414, 2005.
- [8] Thomas Walz, Yoshinori Fujiyoshi, and Andreas Engel. The AQP Structure and Functional Implications. *Handbook of Experimental Pharmacology: Aquaporins*, pages 31–56, 2009.
- [9] Daniel Gorelick, Jeppe Praetorius, Takashi Tsunenari, Soren Nielsen, and Peter Agre. Aquaporin-11: A channel protein lacking apparent transport function expressed in brain. *BMC Biochemistry*, 7(1):14, 2006.
- [10] Tomohiro Itoh, Tatemitsu Rai, Michio Kuwahara, Shigeru B.H. Ko, Shinichi Uchida, Sei Sasaki, and Kenichi Ishibashi. Identification of a novel aquaporin, aqp12, expressed in pancreatic acinar cells. *Biochemical and Biophysical Research Communications*, 330(3):832–838, May 2005.

- [11] Yoshiyuki Morishita, Toshiyuki Matsuzaki, Mariko Hara-chikuma, Ayaka Andoo, Mariko Shimonono, Asako Matsuki, Katsuki Kobayashi, Masahiro Ikeda, Tadashi Yamamoto, Alan Verkman, Eiji Kusano, Shigeo Ookawara, Kuniaki Takata, Sei Sasaki, and Kenichi Ishibashi. Disruption of Aquaporin-11 Produces Polycystic Kidneys following Vacuolization of the Proximal Tubule. *Mol. Cell. Biol.*, 25(17):7770–7779, 2005.
- [12] Ralf Kaldenhoff, Miquel Ribas-Carbo, Jaume Flexas Sans, Claudio Lovisolo, Marlies Heckwolf, and Norbert Uehlein. Aquaporins and plant water balance. *Plant, Cell and Environment*, 31(5):658–666, 2008.
- [13] Urban Johanson, Maria Karlsson, Ingela Johansson, Sofia Gustavsson, Sara Sjoval, Laure Frayse, Alfons R. Weig, and Per Kjellbom. The Complete Set of Genes Encoding Major Intrinsic Proteins in Arabidopsis Provides a Framework for a New Nomenclature for Major Intrinsic Proteins in Plants. *Plant Physiol.*, 126(4):1358–1369, 2001.
- [14] Francoise Quigley, Joshua Rosenberg, Yair Shachar-Hill, and Hans Bohnert. From genome to function: the arabidopsis aquaporins. *Genome Biology*, 3(1):research0001.1–research0001.17, 2001.
- [15] Francois Chaumont, Francois Barrieu, Eva Wojcik, Maarten J. Chrispeels, and Rudolf Jung. Aquaporins Constitute a Large and Highly Divergent Protein Family in Maize. *Plant Physiol.*, 125(3):1206–1215, 2001.
- [16] Junko Sakurai, Fumiyoshi Ishikawa, Tomoya Yamaguchi, Matsuo Uemura, and Masayoshi Maeshima. Identification of 33 Rice Aquaporin Genes and Analysis of Their Expression and Function. *Plant and Cell Physiology*, 46(9):1568–1577, September 2005.
- [17] K B Heller, E C Lin, and T H Wilson. Substrate specificity and transport properties of the glycerol facilitator of Escherichia coli. *J. Bacteriol.*, 144(1):274–278, 1980.
- [18] M. J. Borgnia, D. Kozono, G. Calamita, P. C. Maloney, and P. Agre. Functional reconstitution and characterization of aqpz, the e. coli water channel protein. *J Mol Biol*, 291(5):1169–1179, Sep 1999.
- [19] X. Ding and Y. Kitagawa. Rapid amplification of a water channel-like gene and its flanking sequences from the methanothermobacter marburgensis genome using a single primer PCR strategy. *J Biosci Bioeng*, 92(5):488–491, 2001.
- [20] Mary F Roberts. Osmoadaptation and osmoregulation in archaea: update 2004. *Front Biosci*, 9:1999–2019, Sep 2004.
- [21] John K Lee, David Kozono, Jonathan Remis, Yoshichika Kitagawa, Peter Agre, and Robert M Stroud. Structural basis for conductance by the archaeal aquaporin aqpm at 1.68 Å. *Proc Natl Acad Sci U S A*, 102(52):18932–18937, Dec 2005.

- [22] Nina Pettersson, Caroline Filipsson, Evren Becit, Lars Brive, and Stefan Hohmann. Aquaporins in yeasts and filamentous fungi. *Biol Cell*, 97(7):487–500, Jul 2005.
- [23] Eric Beitz. Aquaporins from pathogenic protozoan parasites: structure, function and potential for chemotherapy. *Biol. Cell*, 97(6):373–383, 2005.
- [24] Eric Beitz. Aquaporin Water and Solute Channels from Malaria Parasites and Other Pathogenic Protozoa. *ChemMedChem*, 1(6):587–592, 2006.
- [25] Martin Hansen, J F J Kun, Joachim E Schultz, and Eric Beitz. A Single, Bi-functional Aquaglyceroporin in Blood-stage *Plasmodium falciparum* Malaria Parasites. *Journal of Biological Chemistry*, 277(7):4874–4882, 2002.
- [26] Slavica Pavlovic-Djuranovic, Joachim E Schultz, and Eric Beitz. A single aquaporin gene encodes a water/glycerol/urea facilitator in *Toxoplasma gondii* with similarity to plant tonoplast intrinsic proteins. *FEBS Lett*, 555(3):500–504, Dec 2003.
- [27] Nestor L. Uzcategui, Alexander Szallies, Slavica Pavlovic-Djuranovic, Monica Palmada, Katherine Figarella, Christoph Boehmer, Florian Lang, Eric Beitz, and Michael Duszenko. Cloning, heterologous expression, and characterization of three aquaglyceroporins from *Trypanosoma brucei*. *Journal of Biological Chemistry*, 279(41):42669–42676, 2004.
- [28] Andrea Montalvetti, Peter Rohloff, and Roberto Docampo. A functional aquaporin co-localizes with the vacuolar proton pyrophosphatase to acidocalcisomes and the contractile vacuole complex of *Trypanosoma cruzi*. *J Biol Chem*, 279(37):38673–38682, Sep 2004.
- [29] Benjamin Gourbal, Niluefer Sonuc, Hiranmoy Bhattacharjee, Danielle Legare, Shyam Sundar, Marc Ouellette, Barry P. Rosen, and Rita Mukhopadhyay. Drug uptake and modulation of drug resistance in leishmania by an aquaglyceroporin. *Journal of Biological Chemistry*, 279(30):31010–31017, 2004.
- [30] Eric Beitz, Dana Becker, Julia Bülow, Christina Conrad, Nadine Fricke, Amornrat Geadkaew, Dawid Krenc, Jie Song, Dorothea Wree, and Binghua Wu. In vitro analysis and modification of aquaporin pore selectivity, 2009.
- [31] Mark L Zeidel, Suresh V Ambudkar, Barbara L Smith, and Peter Agre. Reconstitution of functional water channels in liposomes containing purified red cell CHIP28 protein. *Biochemistry*, 31(33):7436–7440, August 1992.
- [32] T Walz, B L Smith, M L Zeidel, A Engel, and P Agre. Biologically active two-dimensional crystals of aquaporin CHIP. *Journal of Biological Chemistry*, 269(3):1583–1586, 1994.
- [33] P. Pohl, S. M. Saparov, M. J. Borgnia, and P. Agre. Highly selective water channel activity measured by voltage clamp: analysis of planar lipid bilayers reconstituted with purified aqpz. *Proc Natl Acad Sci U S A*, 98(17):9624–9629, Aug 2001.

- [34] Baoxue Yang and A S Verkman. Water and Glycerol Permeabilities of Aquaporins 1-5 and MIP Determined Quantitatively by Expression of Epitope-tagged Constructs in *Xenopus* Oocytes. *Journal of Biological Chemistry*, 272(26):16140–16146, 1997.
- [35] J. S. Jung, R. V. Bhat, G. M. Preston, W. B. Guggino, J. M. Baraban, and P. Agre. Molecular characterization of an aquaporin cDNA from brain: candidate osmoreceptor and regulator of water balance. *Proc Natl Acad Sci U S A*, 91(26):13052–13056, Dec 1994.
- [36] William E C Harries, David Akhavan, Larry J W Miercke, Shahram Khademi, and Robert M Stroud. The channel architecture of aquaporin 0 at a 2.2-Å resolution. *Proc Natl Acad Sci U S A*, 101(39):14045–14050, Sep 2004.
- [37] T. Gonen, P. Sliz, J. Kistler, Y. Cheng, and T. Walz. Aquaporin-0 membrane junctions reveal the structure of a closed water pore. *Nature*, 429:193–197, 2004.
- [38] Tamir Gonen, Yifan Cheng, Piotr Sliz, Yoko Hiroaki, Yoshinori Fujiyoshi, Stephen C. Harrison, and Thomas Walz. Lipid-protein interactions in double-layered two-dimensional aqp0 crystals. *Nature*, 438(7068):633–638, December 2005.
- [39] Richard K Hite, Zongli Li, and Thomas Walz. Principles of membrane protein interactions with annular lipids deduced from aquaporin-0 2d crystals. *EMBO J*, 29(10):1652–1658, May 2010.
- [40] Kazuyoshi Murata, Kaoru Mitsuoka, Teruhisa Hirai, Thomas Walz, Peter Agre, J. Bernard Heymann, Andreas Engel, and Yoshinori Fujiyoshi. Structural determinants of water permeation through aquaporin-1. *Nature*, 407(6804):599–605, October 2000.
- [41] G. Ren, V. S. Reddy, A. Cheng, P. Melnyk, and A. K. Mitra. Visualization of a water-selective pore by electron crystallography in vitreous ice. *Proc Natl Acad Sci U S A*, 98(4):1398–1403, Feb 2001.
- [42] Haixin Sui, Bong-Gyoon Han, John K Lee, Peter Walian, and Bing K Jap. Structural basis of water-specific transport through the AQP1 water channel. *Nature*, 414(6866):872–878, December 2001.
- [43] Joseph D. Ho, Ronald Yeh, Andrew Sandstrom, Ilya Chorny, William E. C. Harries, Rebecca A. Robbins, Larry J. W. Miercke, and Robert M. Stroud. Crystal structure of human aquaporin 4 at 1.8 Å and its mechanism of conductance. *Proceedings of the National Academy of Sciences*, 106(18):7437–7442, 2009.
- [44] Yoko Hiroaki, Kazutoshi Tani, Akiko Kamegawa, Nobuhiko Gyobu, Kouki Nishikawa, Hiroshi Suzuki, Thomas Walz, Sei Sasaki, Kaoru Mitsuoka, Kazushi Kimura, Akira Mizoguchi, and Yoshinori Fujiyoshi. Implications of the aquaporin-4 structure on array formation and cell adhesion. *J Mol Biol*, 355(4):628–639, Jan 2006.

- [45] Kazutoshi Tani, Tadanori Mitsuma, Yoko Hiroaki, Akiko Kamegawa, Kouki Nishikawa, Yukihiro Tanimura, and Yoshinori Fujiyoshi. Mechanism of aquaporin-4's fast and highly selective water conduction and proton exclusion. *J Mol Biol*, 389(4):694–706, Jun 2009.
- [46] Rob Horsefield, Kristina Nordén, Maria Fellert, Anna Backmark, Susanna Tornroth-Horsefield, Anke C. Terwisscha van Scheltinga, Jan Kvassman, Per Kjellbom, Urban Johanson, and Richard Neutze. High-resolution x-ray structure of human aquaporin 5. *Proceedings of the National Academy of Sciences*, 105(36):13327–13332, 2008.
- [47] W. Kukulski, A. D. Schenk, U. Johanson, T. Braun, B. L. de Groot, D. Fotiadis, P. Kjellbom, and A. Engel. The 5a structure of heterologously expressed plant aquaporin sopip2;1. *J Mol Biol*, 350(4):611–616, Jul 2005.
- [48] Susanna Tornroth-Horsefield, Yi Wang, Kristina Hedfalk, Urban Johanson, Maria Karlsson, Emad Tajkhorshid, Richard Neutze, and Per Kjellbom. Structural mechanism of plant aquaporin gating. *Nature*, 439(7077):688–694, February 2006.
- [49] Maria Nyblom, Anna Frick, Yi Wang, Mikael Ekvall, Karin Hallgren, Kristina Hedfalk, Richard Neutze, Emad Tajkhorshid, and Susanna Tornroth-Horsefield. Structural and functional analysis of sopip2;1 mutants adds insight into plant aquaporin gating. *Journal of Molecular Biology*, 387(3):653 – 668, 2009.
- [50] Daxiong Fu, Andrew Libson, Larry J W Miercke, Cindy Weitzman, Peter Nollert, Jolanta Krucinski, and Robert M Stroud. Structure of a Glycerol-Conducting Channel and the Basis for Its Selectivity. *Science*, 290(5491):481–486, 2000.
- [51] Emad Tajkhorshid, Peter Nollert, Morten O Jensen, Larry J W Miercke, Joseph O'Connell, Robert M Stroud, and Klaus Schulten. Control of the Selectivity of the Aquaporin Water Channel Family by Global Orientational Tuning. *Science*, 296(5567):525–530, 2002.
- [52] David F Savage, Pascal F Egea, Yaneth Robles-Colmenares, Joseph D O'Connell, and Robert M Stroud. Architecture and selectivity in aquaporins: 2.5 a x-ray structure of aquaporin z. *PLoS Biol*, 1(3):E72, Dec 2003.
- [53] Jiansheng Jiang, Brenda V. Daniels, and Dax Fu. Crystal Structure of AqpZ Tetramer Reveals Two Distinct Arg-189 Conformations Associated with Water Permeation through the Narrowest Constriction of the Water-conducting Channel. *Journal of Biological Chemistry*, 281(1):454–460, 2006.
- [54] Gerhard Fischer, Urszula Kosinska-Eriksson, Camilo Aponte-Santamaria, Madelene Palmgren, Cecilia Geijer, Kristina Hedfalk, Stefan Hohmann, Bert L. de Groot, Richard Neutze, and Karin Lindkvist-Petersson. Crystal structure of a yeast aquaporin at 1.15 Å reveals a novel gating mechanism. *PLoS Biol*, 7(6):e1000130, 06 2009.

- [55] Zachary E R Newby, Joseph O'Connell III, Yaneth Robles-Colmenares, Shahram Khademi, Larry J Miercke, and Robert M Stroud. Crystal structure of the aquaglyceroporin PfAQP from the malarial parasite *Plasmodium falciparum*. *Nat Struct Mol Biol*, 15(6):619–625, June 2008.
- [56] Jochen S. Hub, Helmut Grubmüller, and Bert L. Groot. Dynamics and energetics of permeation through aquaporins. what do we learn from molecular dynamics simulations? 190:57–76, 2009.
- [57] Bert L de Groot and Helmut Grubmüller. The dynamics and energetics of water permeation and proton exclusion in aquaporins. *Current Opinion in Structural Biology*, 15(2):176–183, 2005.
- [58] Jochen S Hub and Bert L de Groot. Mechanism of selectivity in aquaporins and aquaglyceroporins. *Proceedings of the National Academy of Sciences*, 105(4):1198–1203, 2008.
- [59] Bert L de Groot, Tomaso Frigato, Volkhard Helms, and Helmut Grubmüller. The mechanism of proton exclusion in the aquaporin-1 water channel. *J Mol Biol*, 333(2):279–293, Oct 2003.
- [60] Nilmadhab Chakrabarti, Benoît Roux, and Régis Pomès. Structural determinants of proton blockage in aquaporins. *J Mol Biol*, 343(2):493–510, Oct 2004.
- [61] Nilmadhab Chakrabarti, Emad Tajkhorshid, Benoît Roux, and Régis Pomès. Molecular basis of proton blockage in aquaporins. *Structure*, 12(1):65–74, Jan 2004.
- [62] Jochen S Hub and Bert L de Groot. Does CO₂ Permeate through Aquaporin-1? *Biophysical Journal*, 91(3):842–848, 2006.
- [63] A. S. Verkman. Aquaporins: translating bench research to human disease. *J Exp Biol*, 212(Pt 11):1707–1715, Jun 2009.
- [64] A. S. Verkman. Mammalian aquaporins: diverse physiological roles and potential clinical significance. *Expert Rev Mol Med*, 10:e13, 2008.
- [65] A. S. Verkman. More than just water channels: unexpected cellular roles of aquaporins. *J Cell Sci*, 118(15):3225–3232, 2005.
- [66] P. M. Deen. Requirement of human renal water channel aquaporin-2 for vasopressin-dependent concentration of urine. *Science*, 264:92–95, 1994.
- [67] V. Berry, P. Francis, S. Kaushal, A. Moore, and S. Bhattacharya. Missense mutations in mip underlie autosomal dominant 'polymorphic' and lamellar cataracts linked to 12q. *Nature Genet.*, 25:15–17, 2000.
- [68] A. Shiels and S. Bassnett. Mutations in the founder of the mip gene family underlie cataract development in the mouse. *Nature Genet.*, 12:212–215, 1996.
- [69] S. Steinfeld. Abnormal distribution of aquaporin-5 water channel protein in salivary glands from sjogren's syndrome patients. *Lab. Invest.*, 81:143–148, 2001.

- [70] S. Nielsen. Congestive heart failure in rats is associated with increased expression and targeting of aquaporin-2 water channel in collecting duct. *Proc. Natl Acad. Sci. USA*, 94:5450–5455, 1997.
- [71] R. W. Schrier and P. Y. Martin. Recent advances in the understanding of water metabolism in heart failure. *Adv. Exp. Med. Biol.*, 449:415–426, 1998.
- [72] G. T. Manley. Aquaporin-4 deletion in mice reduces brain edema after acute water intoxication and ischemic stroke. *Nature Med.*, 6:159–163, 2000.
- [73] Zsolt Zador, Shirley Stiver, Vincent Wang, and Geoffrey T. Manley. Role of aquaporin-4 in cerebral edema and stroke. 190:159–170, 2009.
- [74] L. S. King, S. Nielsen, P. Agre, and R. H. Brown. Decreased pulmonary vascular permeability in aquaporin-1-null humans. *Proc. Natl Acad. Sci. USA*, 99:1059–1063, 2002.
- [75] Alan S Verkman. Knock-out models reveal new aquaporin functions. *Handb Exp Pharmacol*, (190):359–381, 2009.
- [76] Mohammad Obaidul Hoque, Jean-Charles Soria, Janghee Woo, Taekyeol Lee, Juna Lee, Se Jin Jang, Sunil Upadhyay, Barry Trink, Constance Monitto, Chantal Desmaze, Li Mao, David Sidransky, and Chulso Moon. Aquaporin 1 is overexpressed in lung cancer and stimulates nih-3t3 cell proliferation and anchorage-independent growth. *Am J Pathol*, 168(4):1345–1353, 2006.
- [77] Norikazu Maeda, Toshiyuki Hibuse, and Tohru Funahashi. Handbook of experimental pharmacology. 190:233–249, 2009.
- [78] Malcolm J Gardner, Neil Hall, Eula Fung, Owen White, Matthew Berriman, Richard W Hyman, Jane M Carlton, Arnab Pain, Karen E Nelson, Sharen Bowman, Ian T Paulsen, Keith James, Jonathan A Eisen, Kim Rutherford, Steven L Salzberg, Alister Craig, Sue Kyes, Man-Suen Chan, Vishvanath Nene, Shamira J Shallom, Bernard Suh, Jeremy Peterson, Sam Angiuoli, Mihaela Pertea, Jonathan Allen, Jeremy Selengut, Daniel Haft, Michael W Mather, Akhil B Vaidya, David M A Martin, Alan H Fairlamb, Martin J Fraunholz, David S Roos, Stuart A Ralph, Geoffrey I McFadden, Leda M Cummings, G Mani Subramanian, Chris Mungall, J Craig Venter, Daniel J Carucci, Stephen L Hoffman, Chris Newbold, Ronald W Davis, Claire M Fraser, and Bart Barrell. Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature*, 419(6906):498–511, October 2002.
- [79] Slavica Pavlovic-Djuranovic, Jürgen F J Kun, Joachim E Schultz, and Eric Beitz. Dihydroxyacetone and methylglyoxal as permeants of the *Plasmodium* aquaglyceroporin inhibit parasite proliferation. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1758(8):1012–1017, 2006.

- [80] Thomas Zeuthen, Binghua Wu, Slavica Pavlovic-Djuranovic, Lars M Holm, Nestor L Uzcategui, Michael Duszenko, Jürgen F J Kun, Joachim E Schultz, and Eric Beitz. Ammonia permeability of the aquaglyceroporins from *plasmodium falciparum*, *toxoplasma gondii* and *trypanosoma brucei*. *Molecular Microbiology*, 61(6):1598–1608, 2006.
- [81] Nathalie Marquis, Benjamin Gourbal, Barry P Rosen, Rita Mukhopadhyay, and Marc Ouellette. Modulation in aquaglyceroporin aqp1 gene transcript levels in drug-resistant leishmania. *Mol Microbiol*, 57(6):1690–1699, Sep 2005.
- [82] Bert L de Groot, D Peter Tieleman, Peter Pohl, and Helmut Grubmüller. Water Permeation through Gramicidin A: Desformylation and the Double Helix: A Molecular Dynamics Study. *Biophysical Journal*, 82(6):2934–2942, 2002.
- [83] Fangqiang Zhu, Emad Tajkhorshid, and Klaus Schulten. Theory and Simulation of Water Permeation in Aquaporin-1. *Biophysical Journal*, 86(1):50–57, 2004.
- [84] Fangqiang Zhu, Emad Tajkhorshid, and Klaus Schulten. Collective Diffusion Model for Water Permeation through Microscopic Channels. *Phys. Rev. Lett.*, 93(22):224501, November 2004.
- [85] Masanori Hashido, Akinori Kidera, and Mitsunori Ikeguchi. Water Transport in Aquaporins: Osmotic Permeability Matrix Analysis of Molecular Dynamics Simulations. *Biophysical Journal*, 93(2):373–385, July 2007.
- [86] E. Matthias Müller, Jochen S Hub, Helmut Grubmüller, and Bert L de Groot. Is tea an inhibitor for human aquaporin-1? *Pflugers Arch*, 456(4):663–669, Jul 2008.
- [87] Jochen S Hub. *Selectivity, Regulation, and Inhibition of Aquaporin Channels. A Molecular Dynamics Study*. PhD thesis, Georg-August-Universität zu Göttingen, 2008.
- [88] Bas W.M. van Balkom, Michelle Boone, Giel Hendriks, Erik-Jan Kamsteeg, Joris H. Robben, H. Christiaan Stronks, Anne van der Voorde, Francois van Herp, Peter van der Sluijs, and Peter M.T. Deen. LIP5 Interacts with Aquaporin 2 and Facilitates Its Lysosomal Degradation. *J Am Soc Nephrol*, 20(5):990–1001, 2009.
- [89] Yangjian Liu, Dominique Promeneur, Aleksandra Rojek, Nirbhay Kumar, Jorgen Fokiaer, Soren Nielsen, Landon S. King, Peter Agre, and Jennifer M. Carbrey. Aquaporin 9 is the major pathway for glycerol uptake by mouse erythrocytes, with implications for malarial virulence. *Proceedings of the National Academy of Sciences*, 104(30):12560–12564, 2007.
- [90] Aleksandra M. Rojek, Mariusz T. Skowronski, Ernst-Martin Füchtbauer, Annette C. Füchtbauer, Robert A. Fenton, Peter Agre, Jorgen Fokiaer, and Soren Nielsen. Defective glycerol metabolism in aquaporin 9 (aqp9) knockout mice. *Proceedings of the National Academy of Sciences*, 104(9):3609–3614, 2007.
- [91] M.P. Allen and D.J. Tildesley. *Computer simulation of liquids*. Clarendon Pr., 1993.

- [92] D. Frenkel and B. Smit. *Understanding molecular simulation: from algorithms to applications*. Computational science. Academic Press, 2002.
- [93] D.C. Rapaport. *The art of molecular dynamics simulation*. Cambridge University Press, 2004.
- [94] Harold A. Scheraga, Mey Khalili, and Adam Liwo. Protein-folding dynamics: Overview of molecular simulation techniques. *Annual Review of Physical Chemistry*, 58(1):57–83, 2007.
- [95] Wilfred F. van Gunsteren, Dirk Bakowies, Riccardo Baron, Indira Chandrasekhar, Markus Christen, Xavier Daura, Peter Gee, Daan P. Geerke, Alice Glättli, Philippe H. Hünenberger, Mika A. Kastenholz, Chris Oostenbrink, Merijn Schenk, Daniel Trzesniak, Nico F. A. van der Vegt, and Haibo B. Yu. Biomolecular modeling: Goals, problems, perspectives. *Angewandte Chemie International Edition*, 45(25):4064–4092, 2006.
- [96] Martin Karplus and J. Andrew McCammon. Molecular dynamics simulations of biomolecules. *Nat Struct Mol Biol*, 9(9):646–652, September 2002.
- [97] D. van der Spoel, E. Lindahl, B. Hess, A. R. van Buuren, E. Apol, P. J. Meulenhoff, D. P. Tieleman, A. L. T. M. Sijbers, K. A. Feenstra, R. van Drunen and H. J. C. Berendsen, *Gromacs User Manual version 4.0*, www.gromacs.org (2005).
- [98] Berk Hess, Carsten Kutzner, David van der Spoel, and Erik Lindahl. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *Journal of Chemical Theory and Computation*, 4(3):435–447, March 2008.
- [99] David Van Der Spoel, Erik Lindahl, Berk Hess, Gerrit Groenhof, Alan E Mark, and Herman J C Berendsen. GROMACS: fast, flexible, and free. *J Comput Chem*, 26(16):1701–1718, December 2005.
- [100] Federick Reif. *Fundamentals of Statistical and Thermal Physics*. Mc.Graw-Hill, 1985.
- [101] William L Jorgensen, David S Maxwell, and Julian Tirado-Rives. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *Journal of the American Chemical Society*, 118(45):11225–11236, January 1996.
- [102] George A Kaminski, Richard A Friesner, Julian Tirado-Rives, and William L Jorgensen. Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides. *The Journal of Physical Chemistry B*, 105(28):6474–6487, July 2001.
- [103] Viktor Hornak, Robert Abel, Asim Okur, Bentley Strockbine, Adrian Roitberg, and Carlos Simmerling. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins*, 65(3):712–725, November 2006.

- [104] Chris Oostenbrink, Alessandra Villa, Alan E. Mark, and Wilfred F. Van Gunsteren. 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(13):1656–1676, 2004.
- [105] B. R. Brooks, C. L. Brooks, A. D. Mackerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York, and M. Karplus. Charmm: The biomolecular simulation program. *J. Comput. Chem.*, 30(10):1545–1614, 2009.
- [106] A. D. MacKerell, D. Bashford, Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, and M. Karplus. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *The Journal of Physical Chemistry B*, 102(18):3586–3616, April 1998.
- [107] R.W. Hockney and J.W. Eastwood. *Computer simulation using particles*. A. Hilger, 1988.
- [108] Loup Verlet. Computer "experiments" on classical fluids. i. thermodynamical properties of lennard-jones molecules. *Phys. Rev.*, 159(1):98, Jul 1967.
- [109] Shuichi Miyamoto and Peter A Kollman. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *Journal of Computational Chemistry*, 13(8):952–962, 1992.
- [110] Berk Hess, Henk Bekker, Herman J C Berendsen, and Johannes G E M Fraaije. LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry*, 18(12):1463–1472, 1997.
- [111] K Anton Feenstra, Berk Hess, and Herman J C Berendsen. Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich systems. *Journal of Computational Chemistry*, 20(8):786–798, 1999.
- [112] Minoru Saito. Molecular dynamics simulations of proteins in solution: Artifacts caused by the cutoff approximation. *The Journal of Chemical Physics*, 101(5):4055–4061, 1994.
- [113] Minoru Saito. Molecular dynamics/free energy study of a protein in solution with all degrees of freedom and long-range coulomb interactions. *The Journal of Physical Chemistry*, 99(46):17043–17048, 1995.
- [114] Ulrich Essmann, Lalith Perera, Max L Berkowitz, Tom Darden, Hsing Lee, and Lee G Pedersen. A smooth particle mesh Ewald method. *The Journal of Chemical Physics*, 103(19):8577–8593, 1995.

- [115] Tom Darden, Darrin York, and Lee Pedersen. Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. *The Journal of Chemical Physics*, 98(12):10089–10092, 1993.
- [116] H J C Berendsen, J P M Postma, W F van Gunsteren, A DiNola, and J R Haak. Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics*, 81(8):3684–3690, 1984.
- [117] Giovanni Bussi, Davide Donadio, and Michele Parrinello. Canonical sampling through velocity rescaling. *The Journal of Chemical Physics*, 126(1):14101, 2007.
- [118] Hans C. Andersen. Molecular dynamics simulations at constant pressure and/or temperature. *The Journal of Chemical Physics*, 72(4):2384–2393, 1980.
- [119] Shuichi Nose. A unified formulation of the constant temperature molecular dynamics methods. *The Journal of Chemical Physics*, 81(1):511–519, 1984.
- [120] Stefan Michael Kast, Kai Nicklas, Hans-Jurgen Bar, and Jurgen Brickmann. Constant temperature molecular dynamics simulations by means of a stochastic collision model. i. noninteracting particles. *The Journal of Chemical Physics*, 100(1):566–576, 1994.
- [121] Stefan Michael Kast and Jurgen Brickmann. Constant temperature molecular dynamics simulations by means of a stochastic collision model. ii. the harmonic oscillator. *The Journal of Chemical Physics*, 104(10):3732–3741, 1996.
- [122] M Parrinello and A Rahman. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics*, 52(12):7182–7190, 1981.
- [123] Schrödinger, LLC. The PyMOL molecular graphics system, version 1.3r1. August 2010.
- [124] William Humphrey, Andrew Dalke, and Klaus Schulten. VMD – Visual Molecular Dynamics. *Journal of Molecular Graphics*, 14:33–38, 1996.
- [125] John G. Kirkwood. Statistical mechanics of fluid mixtures. *The Journal of Chemical Physics*, 3(5):300–313, 1935.
- [126] Guillem Portella Carbo. *Determinants of water and ion permeation through nanopores studied by Molecular Dynamics simulations*. PhD thesis, Georg-August-Universität zu Göttingen, 2008.
- [127] Benoît Roux. The calculation of the potential of mean force using computer simulations. *Computer Physics Communications*, 91(1-3):275 – 282, 1995.
- [128] Shankar Kumar, John M Rosenberg, Djamel Bouzida, Robert H Swendsen, and Peter A Kollman. THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *Journal of Computational Chemistry*, 13(8):1011–1021, 1992.

- [129] Donald Truhlar. *Chipot, C., Pohorille, A., Eds. Free Energy Calculations: Theory and Applications in Chemistry and Biology*, volume 121. Springer Berlin / Heidelberg, 2008-09-01.
- [130] Andrea Amadei, Antonius B. M. Linsen, and Herman J. C. Berendsen. Essential dynamics of proteins. *Proteins: Structure, Function, and Genetics*, 17(4):412–425, 1993.
- [131] Marcus Kubitzki. *Enhanced Conformational Sampling of Proteins Using TEE-REX*. PhD thesis, Georg August Universität zu Göttingen, 2007.
- [132] Kieran Kirk. Channels and transporters as drug targets in the Plasmodium-infected erythrocyte. *Acta Tropica*, 89(3):285–298, 2004.
- [133] Dominique Promeneur, Yangjian Liu, Jorge Maciel, Peter Agre, Landon S King, and Nirbhay Kumar. Aquaglyceroporin PbAQP during intraerythrocytic development of the malaria parasite Plasmodium berghei. *Proceedings of the National Academy of Sciences*, 104(7):2211–2216, 2007.
- [134] Eric Beitz, Slavica Pavlovic-Djuranovic, Masato Yasui, Peter Agre, and Joachim E Schultz. Molecular dissection of water and glycerol permeability of the aquaglyceroporin from Plasmodium falciparum by mutational analysis. *Proceedings of the National Academy of Sciences of the United States of America*, 101(5):1153–1158, 2004.
- [135] H J C Berendsen, J R Grigera, and T P Straatsma. The missing term in effective pair potentials. *The Journal of Physical Chemistry*, 91(24):6269–6271, November 1987.
- [136] G Vriend. WHAT IF: A molecular modeling and drug design program. *Journal of Molecular Graphics*, 8(1):52–56, 1990.
- [137] Maarten G. Wolf, Martin Hoefling, Camilo Aponte-Santamaria, Helmut Grubmueller, and Gerrit Groenhof. g_membed: Efficient insertion of a membrane protein into an equilibrated lipid bilayer with minimal perturbation. *Journal of Computational Chemistry*, 31(11):2169–2174, 2010.
- [138] O Berger, O Edholm, and F Jähnig. Molecular dynamics simulations of a fluid bilayer of dipalmitoylphosphatidylcholine at full hydration, constant pressure, and constant temperature. *Biophys J*, 72(5):2002–2013, May 1997.
- [139] O S Smart, J M Goodfellow, and B A Wallace. The pore dimensions of gramicidin A. *Biophysical Journal*, 65(6):2455–2460, December 1993.
- [140] William L Jorgensen, Jayaraman Chandrasekhar, Jeffrey D Madura, Roger W Impey, and Michael L Klein. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, 79(2):926–935, 1983.
- [141] E M Duffy, D L Severance, and W L Jorgensen. Urea: Potential Functions, log P, and Free Energy of Hydration. *Israel Journal of Chemistry*, 33:323–330, 1993.

- [142] Lorna J Smith, Herman J C Berendsen, and Wilfred F van Gunsteren. Computer Simulation of Urea-Water Mixtures: A Test of Force Field Parameters for Use in Biomolecular Simulation. *The Journal of Physical Chemistry B*, 108(3):1065–1071, January 2004.
- [143] Mark L Zeidel, Soren Nielsen, Barbara L Smith, Suresh V Ambudkar, Arvid B Maunsbach, and Peter Agre. Ultrastructure, Pharmacologic Inhibition, and Transport Selectivity of Aquaporin CHIP in Proteoliposomes. *Biochemistry*, 33(6):1606–1615, February 1994.
- [144] Sapor M Saporov, Satoshi P Tsunoda, and Peter Pohl. Proton exclusion by an aquaglycero-protein: a voltage clamp study. *Biol. Cell*, 97(7):545–550, 2005.
- [145] Fangqiang Zhu, Emad Tajkhorshid, and Klaus Schulten. Pressure-Induced Water Transport in Membrane Channels Studied by Molecular Dynamics. *Biophysical Journal*, 83(1):154–160, 2002.
- [146] Morten Jensen and Ole G Mouritsen. Single-Channel Water Permeabilities of Escherichia coli Aquaporins AqpZ and GlpF. *Biophysical Journal*, 90(7):2270–2284, 2006.
- [147] van Gunsteren W F Hermans J Berendsen HJC Postma JPM. *Intermolecular forces*. Dordrecht, Holland: Reidel, 1981.
- [148] Mario J Borgnia and Peter Agre. Reconstitution and functional comparison of purified GlpF and AqpZ, the glycerol and water channels from Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America*, 98(5):2888–2893, 2001.
- [149] Susanna Tornroth-Horsefield, Kristina Hedfalk, Gerhard Fischer, Karin Lindkvist-Petersson, and Richard Neutze. Structural insights into eukaryotic aquaporin regulation. *FEBS Letters*, 584(12):2580 – 2588, 2010. Gothenburg Special Issue: Molecules of Life.
- [150] An Tanghe, Patrick Van Dijck, and Johan M. Thevelein. Why do microorganisms have aquaporins? *Trends in Microbiology*, 14(2):78 – 85, 2006.
- [151] Christophe Maurel, Thierry Simonneau, and Moira Sutka. The significance of roots as hydraulic rheostats. *Journal of Experimental Botany*, 61(12):3191–3198, 2010.
- [152] Christophe Maurel, Lionel Verdoucq, Doan-Trung Luu, and Vronique Santoni. Plant aquaporins: Membrane channels with multiple integrated functions. *Annual Review of Plant Biology*, 59(1):595–624, 2008.
- [153] Ingela Johansson, Maria Karlsson, Vipula K. Shukla, Maarten J. Chrispeels, Christer Larsson, and Per Kjellbom. Water Transport Activity of the Plasma Membrane Aquaporin PM28A Is Regulated by Phosphorylation. *Plant Cell*, 10(3):451–460, 1998.
- [154] Colette Tournaire-Roux, Moira Sutka, Helene Javot, Elisabeth Gout, Patricia Gerbeau, Doan-Trung Luu, Richard Bligny, and Christophe Maurel. Cytosolic pH regulates root water transport

- during anoxic stress through gating of aquaporins. *Nature*, 425(6956):393–397, September 2003.
- [155] Himanshu Khandelia, Morten A. Jensen, and Ole G. Mouritsen. To gate or not to gate: Using molecular dynamics simulations to morph gated plant aquaporins into constitutively open conformations. *The Journal of Physical Chemistry B*, 113(15):5239–5244, 2009. PMID: 19320451.
- [156] Matthias Fischer and Ralf Kaldenhoff. On the pH regulation of plant aquaporins. *Journal of Biological Chemistry*, 283(49):33889–33892, 2008.
- [157] Y. Duan, C. Wu, S. Chowdhury, M. C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J. Wang, and P. Kollman. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.*, 24(16):1999–2012, 2003.
- [158] Shirley W. I. Siu, Robert Vacha, Pavel Jungwirth, and Rainer A. Bockmann. Biomolecular simulations of membranes: Physical properties from different force fields. *The Journal of Chemical Physics*, 128(12):125103, 2008.
- [159] Nadine Homeyer, Anselm Horn, Harald Lanig, and Heinrich Sticht. Amber force-field parameters for phosphorylated amino acids in different protonation states: phosphoserine, phosphothreonine, phosphotyrosine, and phosphohistidine. *Journal of Molecular Modeling*, 12:281–289, 2006. 10.1007/s00894-005-0028-4.
- [160] Michelle Boone and Peter Deen. Physiology and pathophysiology of the vasopressin-regulated renal water reabsorption. *Pflügers Archiv European Journal of Physiology*, 456:1005–1024, 2008. 10.1007/s00424-008-0498-1.
- [161] Pavel I. Nedvetsky, Grazia Tamma, Sven Beulshausen, Giovanna Valenti, Walter Rosenthal, and Enno Klusmann. Regulation of aquaporin-2 trafficking. 190:133–157, 2009.
- [162] Joseph D. Ho, Ronald Yeh, Andrew Sandstrom, Ilya Chorny, William E. C. Harries, Rebecca A. Robbins, Larry J. W. Miercke, and Robert M. Stroud. Crystal structure of human aquaporin 4 at 1.8 Å and its mechanism of conductance. *Proceedings of the National Academy of Sciences*, 106(18):7437–7442, May 2009.
- [163] D. B. Rubin. The bayesian bootstrap. *The annals of statistics*, 9:130–134, 1981.
- [164] Daniel Seeliger, Jürgen Haas, and Bert L. de Groot. Geometry-based sampling of conformational transitions in proteins. *Structure*, 15(11):1482–1492, November 2007.
- [165] Carol A. Rohl, Charlie E.M. Strauss, Kira M.S. Misura, and David Baker. Protein structure prediction using rosetta. *Methods in Enzymology*, Volume 383:66–93, 2004.

- [166] Jeffrey J. Gray, Stewart Moughon, Chu Wang, Ora Schueler-Furman, Brian Kuhlman, Carol A. Rohl, and David Baker. Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations. *Journal of Molecular Biology*, 331(1):281–299, August 2003.
- [167] Junyu Xiao, Hengchuan Xia, Jiahai Zhou, Ishara F. Azmi, Brian A. Davies, David J. Katzmann, and Zhaohui Xu. Structural basis of vta1 function in the multivesicular body sorting pathway. *Developmental Cell*, 14(1):37 – 49, 2008.
- [168] Andreas Engel, Yoshinori Fujiyoshi, Tamir Gonen, and Thomas Walz. Junction-forming aquaporins. *Current Opinion in Structural Biology*, 18(2):229–235, April 2008.
- [169] Sabine M. Mulders, Gregory M. Preston, Peter M. T. Deen, William B. Guggino, Carel H. van Os, and Peter Agre. Water Channel Properties of Major Intrinsic Protein of Lens. *Journal of Biological Chemistry*, 270(15):9010–9016, 1995.
- [170] Karin L. Németh-Cahalan and James E. Hall. pH and calcium regulate the water permeability of aquaporin 0. *Journal of Biological Chemistry*, 275(10):6777–6782, 2000.
- [171] Richard Hite, Tamir Gonen, Stephen Harrison, and Thomas Walz. Interactions of lipids with aquaporin-0 and other membrane proteins. *Pflügers Archiv European Journal of Physiology*, 456(4):651–661, 2008-07-01.
- [172] Teruhisa Hirai, Kaoru Mitsuoka, Akinori Kidera, and Yoshinori Fujiyoshi. Simulation of charge effects on density maps obtained by high-resolution electron crystallography. *Journal of Electron Microscopy*, 56(4):131–140, 2007.
- [173] Matthias Rarey, Bernd Kramer, Thomas Lengauer, and Gerhard Klebe. A fast flexible docking method using an incremental construction algorithm. *Journal of Molecular Biology*, 261(3):470–489, August 1996.
- [174] Junmei Wang, Wei Wang, Peter A Kollman, and David A Case. Automatic atom type and bond type perception in molecular mechanical calculations. *J Mol Graph Model*, 25(2):247–260, October 2006.
- [175] Junmei Wang, Romain M Wolf, James W Caldwell, Peter A Kollman, and David A Case. Development and testing of a general amber force field. *J Comput Chem*, 25(9):1157–1174, July 2004.

CURRICULUM VITAE: Camilo Andrés Aponte-Santamaría

Education

2007-2010 **Doctor in Natural Sciences**, Max Planck Institute for Biophysical Chemistry, Göttingen (Germany). Supervisor: Prof. Dr. Bert de Groot. Thesis project: *Understanding the molecular machinery of aquaporins through molecular dynamics simulations*

2005-2007 **Magister Science on Physics**, National University of Colombia, Bogotá (Colombia). Supervisor: Prof. José Daniel Muñoz. Thesis: *Brownian Dynamics simulations of the rotation of the F_1 -ATPase molecular motor*.

2000-2005 **Physics**, National University of Colombia, Bogotá (Colombia). Supervisors: Prof. Ramón Fayad and Prof. José Daniel Muñoz. Thesis: *The Gramicidin A Channel: A comparison between an elemental permeation model and a Brownian Dynamics simulation*.

1994-1999 **High school**. Colegio CAFAM, Bogotá (Colombia).

Articles in peer-reviewed journals

- J. S. Hub, C. Aponte-Santamaria, H. Grubmüller and B. L. de Groot. *Voltage-regulated water flux through aquaporin channels in silico*. Biophys. J. (letters). 99:L97-L99 (2010).
- C. Aponte-Santamaria, J. S. Hub and B. L. de Groot. *Dynamics and energetics of solute permeation through the Plasmodium falciparum aquaglyceroporin*. Phys. Chem. Chem. Phys. 12:10246-10254 (2010).

- M. Wolf , M. Hoefling , C. Aponte-Santamaria, H. Grubmüller , and G. Groenhof. *g_membed: Efficient insertion of a membrane protein into an equilibrated lipid bilayer with minimal perturbation*. J. Comp.Chem. 31: 2169-2174 (2010).
- G. Fischer, U. Kosinska-Eriksson, C. Aponte-Santamaria, M. Palmgren, C. Geijer, K. Hedfalk, S. Hohmann, B. L. de Groot, R. Neutze, K. Lindkvist-Petersson. *Crystal Structure of a Yeast Aquaporin at 1.15 Å Reveals a Novel Gating Mechanism*. PLoS Biology. 7: e1000130 (2009).
- C. Aponte-Santamaría, J. D. Muñoz, R. F. and M. Camacho. *Effect of the dielectric media on the conductance of the Gramicidin A channel studied by brownian dynamics simulations*. Rev. Col. Fís. In spanish. 40(2), 466 (2008).
- C. Aponte-Santamaría, J. D. Muñoz and R. Fayad. *Brownian Dynamics Simulation of the Ionic Transport Through the Gramicidin A Channel*. Rev. Col. Fís. In spanish. 38(4), 1639 (2006).

Oral conference contributions

- *Dynamics and energetics of solute permeation through the Plasmodium falciparum malaria aquaglyceroporin*. International Workshop on Epithelial Transport, 06-2010, Strobl (Austria).
- *The gating mechanism of Yeast Aquaporin studied by molecular dynamics simulations*
 - 54th Annual Meeting of the Biophysical Society, 02-2010, San Francisco (USA).
 - International Workshop From Structure to Function: Influx and Efflux Systems, 05-2009, Cagliari (Italy).
 - Workshop on Computer Simulation and Theory of Macromolecules, 04-2009, Huenfeld (Germany).