

**Selectivity, Regulation, and Inhibition of  
Aquaporin Channels.  
A Molecular Dynamics Study**

Dissertation  
zur Erlangung des Doktorgrades  
der mathematisch-naturwissenschaftlichen Fakultäten  
der Georg-August-Universität zu Göttingen

Vorgelegt von  
**Jochen Sebastian Hub**  
aus Stuttgart

Göttingen 2008

**D 7**

**Referent: Prof. Dr. Helmut Grubmüller**

**Koreferent: Prof. Dr. Tim Salditt**

**Tag der mündlichen Prüfung: 28. Januar 2008**





## Abstract

Aquaporins (AQPs) form a large family of transmembrane protein channels which have been found throughout nature. AQPs facilitate the highly efficient and selective flux of water across biological membranes, whereas related aquaglyceroporins are additionally permeated by small organic compounds such as glycerol or urea. For the present thesis we employed molecular dynamics simulations to study the selectivity, inhibition, as well as a putative regulatory mechanism of aquaporins. At first, the permeation of apolar gas molecules such as carbon dioxide through aquaporin-1 (AQP1) is inspected. The simulations reveal that the permeation of such gases through AQP1 can only be inspected in membranes with unusually low intrinsic gas permeability. In the following, a typical member of the water-conducting AQP subfamily (AQP1) is compared to a member of the aquaglyceroporin subfamily (GlpF). It is found that AQP1 forms a filter which allows the permeation of small polar solutes only, whereas GlpF is less selective allowing the permeation of all considered solutes except for urea. Water-protein interactions complemented by steric restraints are shown to play a key role in the molecular mechanism of selectivity. The putative regulation of the water flux through AQP1 by an electrostatic membrane potential is addressed. In addition, through a number of computational approaches the putative inhibition of AQP1 by tetraethylammonium (TEA) is investigated. TEA is shown to bind into the channel entrance while inhibiting the water flux partially. Finally, we present an extensive analysis of short-range order and collective dynamics of a lipid bilayer membrane. The simulations are compared to scattering experiments. It is shown how the interpretation of scattering data can be aided by simulations.





## Zusammenfassung

Aquaporine (AQPs) bilden eine große Proteinfamilie von Transmembrankanälen, die in der gesamten belebten Natur gefunden wurden. AQPs ermöglichen einen schnellen und selektiven Fluss von Wasser über biologische Membranen. Die Unterfamilie der Aquaglyceroporine ermöglicht zusätzlich die Permeation verschiedener organischer Verbindungen wie Glycerin oder Harnstoff. Für diese Dissertation wurden AQPs mit Hilfe von Molekulardynamiksimulationen untersucht, insbesondere im Hinblick auf Selektivität, Inhibition sowie auf einen möglichen Regulationsmechanismus. Zunächst beschäftigt sich die Dissertation mit der Permeation unpolarer Gase wie Kohlenstoffdioxid durch Aquaporin-1 (AQP1). Die Simulationen zeigen, dass die Permeation solcher Gase durch AQP1 höchstens in Membranen mit ungewöhnlich geringer intrinsischer Gaspermeabilität zu erwarten ist. Im Folgenden werden je ein Vertreter der Unterfamilien der wasserleitenden Aquaporine (AQP1) und der Aquaglyceroporine (GlpF) verglichen. Es wird gezeigt, dass AQP1 einen Filter bildet, der die Permeation von kleinen polaren Molekülen erlaubt, während GlpF die Permeation aller betrachteter Moleküle außer Harnstoff ermöglicht. Wasser-Protein-Wechselwirkungen, ergänzt durch sterische Beschränkungen, spielen eine Schlüsselrolle für den molekularen Selektivitätsmechanismus der Kanäle. Eine mutmaßliche Spannungsregulation des Wasserflusses durch AQP1 wird diskutiert. Das folgende Kapitel verwendet eine Kombination von computergestützten Verfahren, um die mögliche Inhibition von AQP1 durch Tetraethylammonium (TEA) zu untersuchen. Es wird gezeigt, wie TEA in den Kanaleingang bindet und den Wasserfluss teilweise inhibiert. Schließlich präsentieren wir eine umfangreiche Analyse der kurzreichweitigen Ordnung und kollektiven Dynamik in Bilipidmembranen. Die Simulationen werden mit Streuexperimenten verglichen. Es wird gezeigt, wie Simulationen die Interpretation von Streudaten erleichtern können.



# Contents

<b>Abstract</b>	<b>iii</b>
<b>Zusammenfassung</b>	<b>v</b>
<b>1. Introduction</b>	<b>5</b>
<b>2. Theory and Methods</b>	<b>15</b>
2.1. Principles of molecular dynamics simulations . . . . .	15
2.1.1. The three approximations of MD . . . . .	15
2.1.2. Simulation details . . . . .	18
2.2. Potentials of mean force . . . . .	21
2.2.1. Definition . . . . .	21
2.2.2. The PMF is a free energy profile . . . . .	22
2.2.3. Simulation methods to determine the PMF . . . . .	23
2.2.4. Umbrella sampling simulations . . . . .	26
2.2.5. Weighted histogram analysis method . . . . .	28
2.3. Validation of solute parameters . . . . .	31
<b>3. Gas permeation through Aquaporin-1</b>	<b>33</b>
3.1. Does CO <sub>2</sub> permeate through Aquaporin-1? . . . . .	34
3.1.1. Motivation . . . . .	34
3.1.2. Methods . . . . .	35
3.1.3. Results and Discussion . . . . .	40
3.1.4. Conclusions . . . . .	47
3.2. O <sub>2</sub> permeation through Aquaporin-1 . . . . .	47
<b>4. The selectivity mechanism of aquaporins and aquaglyceroporins</b>	<b>51</b>
4.1. Introduction . . . . .	51

4.2. Methods . . . . .	53
4.2.1. Simulation setup . . . . .	53
4.2.2. Construction of PMFs . . . . .	55
4.3. Results . . . . .	59
4.3.1. Solute permeabilities . . . . .	59
4.3.2. Solute hydrophobicity . . . . .	63
4.3.3. Water–channel interactions as selectivity mechanism . . . . .	66
4.4. Discussion . . . . .	67
4.5. Conclusions . . . . .	68
<b>5. Is aquaporin-1 a voltage-regulated water channel?</b>	<b>69</b>
5.1. Methods . . . . .	70
5.2. Results and discussion . . . . .	71
5.2.1. Single channel water permeability $p_f$ . . . . .	71
5.2.2. A putative regulatory mechanism . . . . .	73
5.3. Conclusion . . . . .	75
<b>6. Inhibition of aquaporin-1</b>	<b>77</b>
6.1. Introduction — Aquaporins as novel drug targets . . . . .	77
6.2. Inhibition of aquaporin-1 by tetraethylammonium . . . . .	79
6.2.1. Methods . . . . .	79
6.2.2. Binding site . . . . .	82
6.2.3. Binding affinity and $IC_{50}$ estimates . . . . .	85
6.2.4. Inhibitory effect of TEA on hAQP1 . . . . .	86
6.3. Summary . . . . .	87
<b>7. Short range order and collective dynamics of lipid bilayers</b>	<b>89</b>
7.1. Motivation . . . . .	89
7.2. Methods . . . . .	91
7.3. Electron density . . . . .	95
7.4. Static structure factors of lipid bilayers . . . . .	96
7.4.1. Density correlations in the membrane plane . . . . .	96
7.4.2. Including correlations perpendicular to the membrane: reciprocal space mapping . . . . .	99
7.5. Radial distribution functions . . . . .	101



---

7.5.1. Head group structure . . . . .	101
7.5.2. Radial distribution function of lipid tails . . . . .	103
7.6. Lipid tail packing and area per lipid . . . . .	103
7.6.1. Invariance of RDF and structure factor as a function of the area per lipid . . . . .	104
7.6.2. Linear dependence between correlation length and area per lipid . . . . .	106
7.6.3. Determination of structural parameters from the experimen- tal structure factor . . . . .	107
7.7. Dynamic structure factor . . . . .	108
7.7.1. Definitions and theory . . . . .	108
7.7.2. Comparison to experiment . . . . .	110
7.7.3. Does the three-effective-eigenmode model hold? . . . . .	112
7.8. Conclusion . . . . .	113
<b>8. Summary and conclusions</b>	<b>115</b>
<b>A. Appendix</b>	<b>121</b>
<b>Bibliography</b>	<b>123</b>



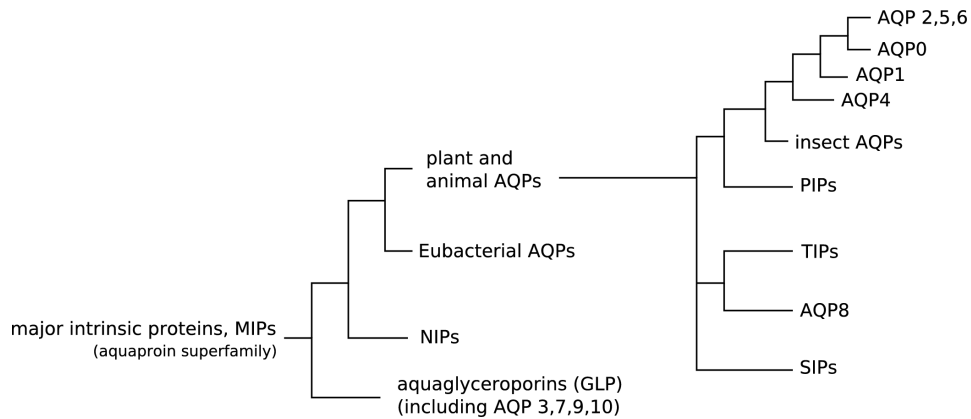


# 1. Introduction

Water is the medium of life. Its unique properties such as the high evaporation temperature combined with a large heat capacity renders water an ideal solvent for biological processes. From what is known today, life is unimaginable without water.

A fundamental strategy adopted by any living organism is to separate the inside of the organism from the outside. At an early state of the evolution, this concept has been realized by the formation of cells which are separated from their environment by a biological membrane. To allow tight control on the biological processes, eukaryotic cells are additionally subdivided into compartments which are again separated from by membranes from the surrounding the cell plasma.

Since water cannot diffuse freely across biological membranes, water channels need to be incorporated into the membranes if a substantial water flux is required. In fact, the existence of intrinsic membrane water channels has been postulated several decades ago. In the 1960s it was found that cells in the amphibian skin can easily be permeated by water (1). Likewise, red blood cell membranes were known to show large water permeability which can be reversibly inhibited by mercury (2, 3, 4). However, such channels remained unidentified until 1992, when *Xenopus* oocyte membranes were observed to exhibit an increased water permeability after injection of cRNA of the so called CHIP28 protein, a membrane protein whose function was unknown in these days (5). (CHIP28 stood for CHannel-forming Integral Protein of 28 kDa.) Today, CHIP28 is called aquaporin-1 (AQP1) and represents the first member of a large protein family termed *major intrinsic proteins* (MIP) whose function could be identified. In mammals, AQP1 turned out to be expressed in a number of different tissues such as red blood cells, kidney, lung, brain, capillaries, and the eye (6). Members of the MIP family have been discovered throughout the biosphere, in organisms ranging from archae and bacteria to fungi, plants, insects and mammals. In 2005, approximately 450 different and (almost) complete amino

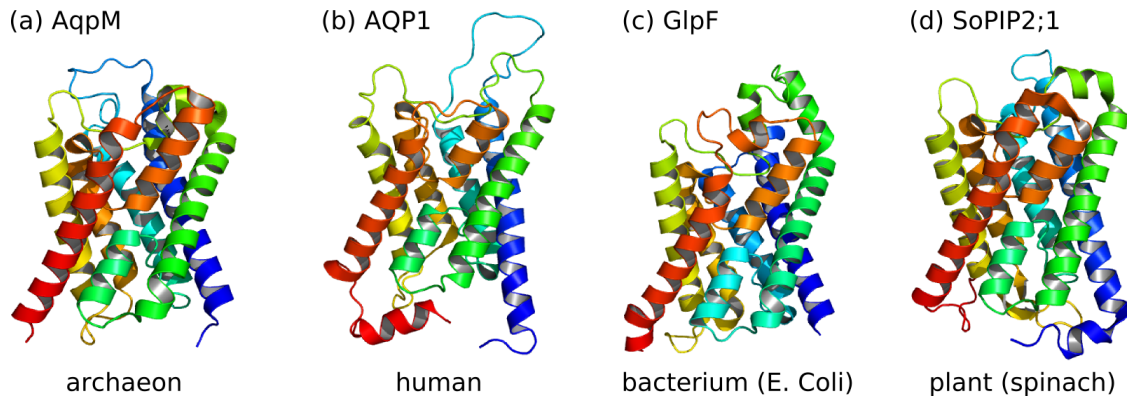


**Figure 1.1.:** Simplified evolutionary tree of the aquaporin superfamily (also termed MIPs). For more detailed phylogenetic trees we refer to Zardoya (7).

acid sequences were listed in sequence databases (7). All MIPs are homologous, indicating that they all may descend from an ancestral aquaporin prototype. Hence, MIPs have probably appeared in a very early state of evolution, which highlights the importance of water flux across biological membranes.

The MIP family can be divided in two subfamilies, in aquaporins (AQP) which are generally considered as pure water channels, and aquaglyceroporins (GLPs) which are additionally permeated by solutes that do not diffuse sufficiently rapid across membranes, such as glycerol, urea, or ammonia. Unfortunately, the nomenclature is ambiguous. In many publications all MIPs are referred to as aquaporins and aquaglyceroporins are rather considered as a subgroup of an aquaporin superfamily. Applying phylogenetic analysis, the ordinary AQPs can be further divided into subgroups, including eubacterial AQPs, animal AQPs, and plant aquaporins (7). The subfamilies of plant MIPs are referred to as NIPs (Nod26-like intrinsic proteins), TIPs (tonoplast intrinsic protein), PIPs (plasma membrane intrinsic proteins), and SIPs (small-basic intrinsic proteins). A simplified scheme of the evolutionary relationships within the AQP superfamily is shown in figure 1.1.

So far, 13 different aquaporins have been discovered in humans. They are specified as AQP0 to AQP12 and have been found in numerous tissues including red blood cells, leukocytes, kidney, liver, lung, brain, eye, testis, several glands, stomach, intestine, blood capillaries, and skin. AQP1, 2, 4, 5, and 8 are selective and efficient water channels. AQP3, 7, 9, and 10 are additionally permeated by urea and glycerol, AQP7 and 9 as well by arsenite (6). AQP6 is restricted to intracellu-

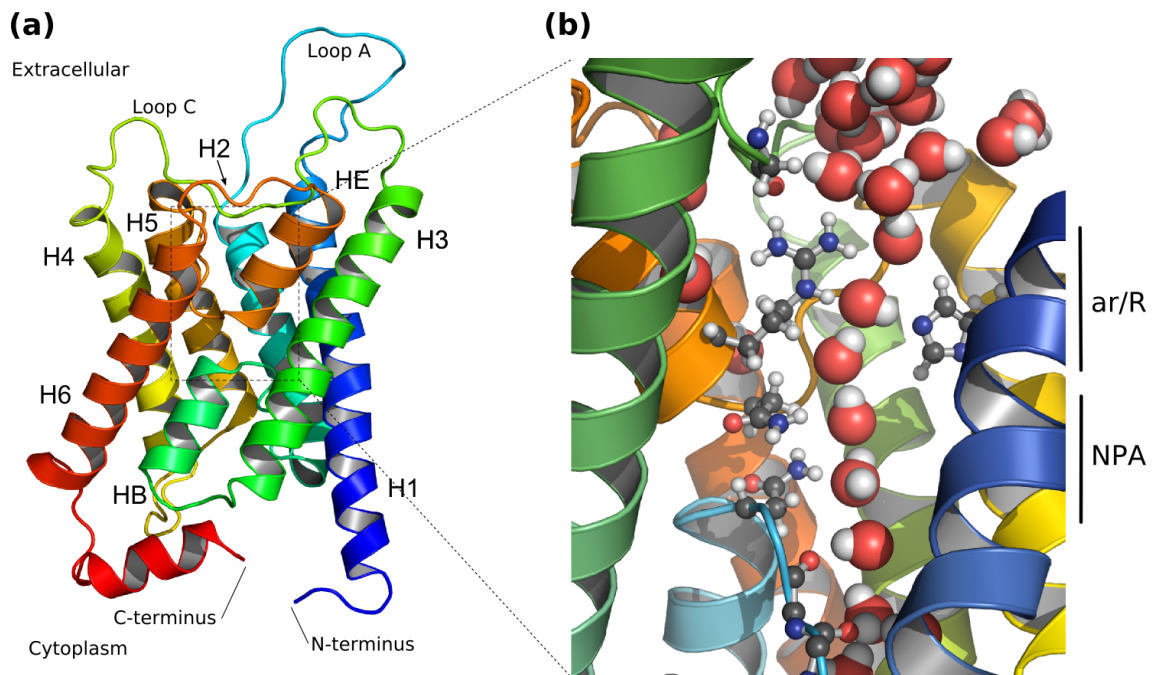


**Figure 1.2.:** Crystal structure of aquaporins from four highly distinct organisms, archael AqpM (a), human AQP1 (b), bacterial GlpF (c), and SoPIP2;1 from spinach (13, 14, 15, 16). All aquaporins share a similar structure including six transmembrane helices and two half helices (compare fig. 1.3a).

lar vesicles, allows the passage of anions such as  $\text{NO}_3^-$  or  $\text{Cl}^-$ , and can be rapidly gated by pH (8). AQP0 is only a poor water channel, but it is known to play an important role in the regular stacking of fiber cells in the eye lens, since mutations in AQP0 were reported to cause cataract (9, 10). The functions of AQP11 and AQP12 are still unknown (11, 12). For more extensive discussions on the diverse functions of aquaporins we refer to the review by King *et al.* (6).

**The unique aquaporin structure** By 2007, the structures of seven distinct aquaporins have been resolved at high resolution by x-ray or electron crystallography (13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24). Remarkably, all aquaporins share a common fold. This is highlighted in figure 1.2 which displays aquaporins from four highly distinct species that have been separated at an early state of the evolution.

The common structural features of aquaporins are depicted in figure 1.3a. Aquaporins are composed of six transmembrane helices (H1-H6) and two half-helices, denoted HB and HE. The loops connecting the helices may significantly differ in length and sequence among different AQPs. Aquaporins have evolved via gene duplication (25). Each repeat contains three transmembrane helices and one half helix. The two repeats are oriented by  $180^\circ$  to each other. The N-terminal segment H1-H2-HB-H3 shares approximately 20% sequence identity with the C-terminal segment H4-H5-HE-H6 (26). Noteworthy, aquaporins form tetramers under physi-



**Figure 1.3.:** (a) Structure of aquaporin-1 (AQP1) as an example for the common fold of aquaporins. AQPs are composed of six transmembrane helices, denoted H1-H6, and two half-helices HE and HB. Loops A and C significantly differ in sequence and length within the aquaporin family (compare fig. 1.2). (b) Snapshot of a molecular dynamics simulation showing a single file of water inside the AQP1 pore. Some water-interacting amino acid side chains are shown in ball-and-stick representation. As indicated by the black bar, the two conserved Asn-Pro-Ala (NPA) motifs are located at the end of the two half-helices. The asparagines of the NPA motifs form strong hydrogen bonds to permeating water molecules. Closer to the extracellular exit of the channel the aromatic/arginine constriction region (ar/R) forms the narrowest part of the channel (15).

ological conditions, with each monomer contributing one functional water pore. An overview on AQP crystal structures and current structural research on AQPs can be found in Gonen *et al.* (27).

**Aquaporin function** Water permeates in a single-file manner across AQP channels as shown in fig. 1.3b. The ordered motion of water molecules is ensured by the regular arrangement of hydrogen bond donors and acceptors along the channel pore (15). Hydrophilic groups are located at one side of the pore, forming a lattice of potential hydrogen bond partners for water molecules. The channel pocket opposite to the H-bond partners is hydrophobic, which ensures a well-ordered configuration of water molecules along their H-bond partners. This structure allows



rapid permeation of up to  $3 \times 10^9$  water molecules per second and channel (28, 29).

Two key interaction sites for permeating water molecules have been identified in AQPs (30). Two conserved Asn-Pro-Ala (NPA) motifs are located at the inner-pore ends of the two half-helices HB and HE (compare fig. 1.3b). The two asparagines of the NPA site point into the pore and form strong hydrogen bonds to passing water molecules. The dipoles of the two half helices generate a large electrostatic potential around the NPA site which makes dipoles of the water molecules rotate by  $\sim 180^\circ$  upon permeation through the channel (30).  $10 \text{ \AA}$  closer to the extracellular exit of the pore, the aromatic/arginine (ar/R) constriction region forms a second key interaction site (fig. 1.3b). Here, a conserved arginine together with aromatic residues constitute the narrowest part of the channel (15).

Protons (or hydronium ions) are strictly excluded from AQP channels. This feature is pivotal for the cell since it maintains the pH gradient across the membrane. The pH gradient is an energy source which is employed by the enzyme ATPase to synthesize ATP, the energy unit of the cell (31, 32). The fact that AQPs exclude protons is remarkable, since other single-file water channels such as bacteriorhodopsin (33) and gramicidin (34, 35) facilitate rapid proton translocation by so-called the *Grotthuss* mechanism (36), the “hop” of protons from an hydronium ion to a neighboring water molecule along a hydrogen bond (33, 37). While early studies focused on a discontinuous H-bond network inside the pore to explain the proton exclusion (30, 38), more recent studies highlight the large energetic barrier that hydronium ions encounter near the NPA region (39, 40, 41, 42, 43, 44). Whether the barrier is of electrostatic origin or due to desolvation effects is still a matter of debate (45, 46).

**Malfunctions of aquaporins cause diseases** Since the discovery of AQPs medical implications of AQPs have drawn increasing attention. Malfunctions of AQPs may cause a number of pathological conditions which are often associated to misregulation of water flux. Examples are the eye disease *glaucoma* which can result in blindness (47), or *diabetes insipidus* characterized by the inability of the kidney to concentrate the urine (48). Moreover, AQPs have been shown to be associated to life threatening conditions such as brain edema and tumor growth (49, 50, 51, 52, 53, 54, 55). These examples highlight the impact of AQP research far beyond basic research and the need to understand the physiological functions of AQPs in

the human body. For a more detailed discussion of medical implications of AQP1 we refer to chapter 6.1 and a number of reviews (56, 57, 58).

**Physiological functions beyond water transport** Within the last years substantial progress has been made in understanding the manifold physiological functions of AQPs and —on a biophysical level— the mechanisms of water permeation and proton exclusion. Current research focuses increasingly on permeation characteristics of AQPs beyond their function as mere water channels. For example, there has been a controversial discussion on putative gas permeation through AQP1 ever since AQP1 was reported to increase CO<sub>2</sub> flux across oocyte membranes (59). Ammonia transport through AQPs is an active field of research (60, 61) and glycerol conduction through aquaporins has been shown to be physiologically relevant (62). Another interesting development in AQP research is the focus on gating mechanisms which have been discovered in plant AQPs and human AQP6 (8, 63). Those recent developments are the starting point for the present thesis.

**Experimental techniques. Why simulations?** Water permeation experiments are often performed by injecting cRNA of an aquaporin channel into *Xenopus* oocytes (frog eggs) (5) or by directly reconstituting the AQP protein into liposomes (29). Subsequently, the swelling of the oocyte or liposome in response to an osmotic shock is measured. The permeation of solutes such as glycerol can be determined by controlling the osmolality of the suspension with a nonpermeable solute such as sugar (64).

Permeation of CO<sub>2</sub> through AQP1 is more challenging as it is measured via acidification with the help of carbonic anhydrase, an enzyme which catalyzes the reaction of CO<sub>2</sub> to carbonic acid (59). Unfortunately, it is not straightforward to distinguish whether the CO<sub>2</sub> permeates through the channel or across the membrane. Although widely used, the inhibition of AQP1 by Hg<sup>2+</sup> is not specific as Hg<sup>2+</sup> can affect the membrane or inhibit other proteins. Moreover, indirect effects of the protein and the experimental procedure on permeation across the membrane cannot be excluded (65). The physiological role of AQPs is often investigated by knockout or overexpression of the corresponding AQP gene. The interpretation of such experiments is not trivial since cells may compensate the knockout of an AQP channel by overexpression of a related channel (66) or other side effects (67).

Since computer simulations allow to control solute pathways they are a useful



tool to complement permeation experiments. Additionally, having all atom coordinates and interactions at hand, they provide insights into molecular mechanisms which drive biological processes such as solute permeation. The main drawbacks of simulations, i.e. limited sampling and force field limitations, must be alleviated by appropriate simulation techniques and extensive comparison to experimental data.

## Aim of the present thesis

The present thesis aims towards a deeper understanding of the mechanisms that govern permeation through AQPs. In a first step, we investigate the role of AQP1 in CO<sub>2</sub> permeation across biological membranes (chapter 3). Chapter 4 describes a general mechanism underlying the selectivity for various solutes permeating through aquaporins and aquaglyceroporins. Chapter 5 reports on a putative regulation of water flux through AQP1 by an electrostatic membrane potential, and chapter 6 addresses the inhibitory effect of tetraethylammonium on AQP1 mediated water flux, a contribution to the search for highly desired inhibitors for AQP1. The final chapter 7 focuses on structural and dynamic properties of the medium surrounding the AQP, namely the lipid bilayer. Each chapter provides additional introductory background.

**Gas permeation through AQP1** Gas transport is pivotal for any living organism. For example, the transport of dioxygen O<sub>2</sub> from the lung to virtually any tissue of the human body must be maintained to allow the burning of sugar. In turn, the metabolism requires the evacuation of CO<sub>2</sub> via the lung. Along their way the gas molecules cross a number of biological membranes such as membranes of the pulmonary alveoli or the red blood cell membranes. Until recently, it was generally accepted that gas molecules can freely diffuse across biological membranes.

However, in 1998 it has been reported that AQP1 facilitates CO<sub>2</sub> flux across oocyte membranes (59). The publication triggered a lively discussion whether AQP1 is permeated by CO<sub>2</sub> and whether CO<sub>2</sub> transport by AQP1 has any physiological relevance. Several publications emphasized the significance of CO<sub>2</sub> permeation through AQPs and their physiological role for CO<sub>2</sub> permeation across membranes in humans and plants (68, 69, 70). Other studies did not observe any CO<sub>2</sub> permeation through AQPs and questioned any physiological relevance (65, 71).

Chapter 3 of the current thesis tries to resolve the controversy. We employed molecular dynamics (MD) simulations to study the energetics of CO<sub>2</sub> permeation along all possible pathways for a CO<sub>2</sub> molecule across an AQP1 tetramer embedded in a model membrane. The results suggest a physiological role of AQP1 as CO<sub>2</sub> channel *only* in membranes with unusually low intrinsic CO<sub>2</sub> permeability.

**The selectivity mechanism** Numerous permeation experiments revealed diverse permeation characteristics of aquaporins and aquaglyceroporins (72). AQPs were reported to be permeated by water, CO<sub>2</sub>, nitric oxide NO, glycerol, urea, ammonia, O<sub>2</sub>, hydrogen peroxide H<sub>2</sub>O<sub>2</sub>, arsenite, antimonite, and silicon (6, 60, 61, 73, 74, 75, 76, 77, 78, 79). However, a unifying picture which explains the diverse properties of aquaporins has not evolved yet. In addition, the measured effects are sometimes small and it is not clear whether all experiments do measure solute flux through the AQP pore or whether some results are governed by indirect effects (65).

Therefore, chapter 4 takes a step beyond chapter 3 and employs extensive MD simulations to elucidate a general selectivity mechanism underlying the characteristics of aquaporins and aquaglyceroporins. Moreover, the chapter aids the interpretation of permeation experiments by comparing the permeation pathway through the AQP pores with the alternative pathway across two model membranes.

**Voltage-regulated water flux?** The tight regulation of water permeation across membranes is essential for any living cell. By regulating the water permeability of its membranes the cell can react to changes in the availability of water. Many plants, for example, need to cope with wet conditions as well as longer periods of drought. Several mechanisms to regulate the water flux via AQPs have been discovered during the last years. The density of AQPs in the plasma membrane can be regulated by hormones which trigger the trafficking of AQPs from storage vesicles inside the cell to the plasma membrane (80, 81, 82, 83, 84). Mammalian AQP0, 3, and 6 as well as a PIP channel in plant roots were found to be sensitive to pH or Ca<sup>2+</sup> concentration (8, 16, 63, 85, 86, 87, 88).

In contrast to many ion channels (89, 90), AQPs have not been reported to be sensitive to an electrostatic membrane potential. However, we observed a remarkable correlation between the membrane potential and the permeability of AQP1 during MD simulations, which is further outlined in chapter 5.





**The search for inhibitors** Effective and selective inhibitors for AQPs are highly desired. AQP1 is expressed in vascular membranes of tumor tissues (49). In mice lacking AQP1, tumors showed impaired growth of blood capillaries (angiogenesis), leading to reduced tumor growth (50, 51). Hence, effective inhibitors for AQP1 may turn out as new antiangiogenic drugs. Other conditions which could be treated by AQP inhibitors include brain edema, glaucoma, or renal disorders.

Unfortunately, selective and effective inhibitors for AQPs remain unknown. In chapter 6 we investigate the binding site and the moderate inhibition of AQP1 by tetraethylammonium (TEA). TEA does not selectively bind to AQP1 nor does it inhibit the water flux completely, but it may turn out as a lead compound for AQP1 inhibition.

**Structure and dynamics of lipid membranes** MD simulations are based on a force field model and must be compared continuously to experimental data. The comparisons are essential to both validate the methodology and to indicate the limits of present force fields which can, in turn, initiate further force field improvements. The present thesis aims to contribute to this process by considering a number of structural quantities of the medium surrounding the AQP, i.e. the lipid bilayer membrane. Therefore, chapter 7 presents comparisons between simulation and experiments in a novel way, with a focus on the *lateral* (parallel to the membrane) ordering of the lipid molecules. In addition, it is demonstrated how simulations help to relate quantities in reciprocal space which can be measured experimentally to real-space properties such as the area per lipid or the tail-tail distance of the lipids. The mapping from reciprocal space to real space quantities represents a good example how the interpretation of experimental results can be aided by simulations.



## 2. Theory and Methods

This chapter introduces the methodological framework that was applied in this thesis. The results have been derived from molecular dynamics (MD) simulation that are briefly discussed in the following section. More details on the applied methods are presented in the corresponding section of chapters 3–7.

### 2.1. Principles of molecular dynamics simulations

MD simulations model the motion of molecular systems by numerically integrating Newton’s equations of motion. Nowadays, MD simulations of small proteins cover timescales of a few microseconds. On shorter timescales, systems of up to  $\sim 1$  million atoms can be described.

#### 2.1.1. The three approximations of MD

Conventional MD simulations are based on three approximations: (i) the decoupling of the motion of nuclei and electrons (Born-Oppenheimer approximation), (ii) the assumptions that the nuclei motions can be described classically, and (iii) the application of empirical force fields. A detailed discussion on the hierarchical approximations of MD can be found in reference (91).

**(i) Born-Oppenheimer approximation** The time evolution of a molecular system is described by the time-dependent Schrödinger equation

$$\mathcal{H}\psi = i\hbar\frac{\partial\psi}{\partial t}, \quad (2.1)$$

where  $\mathcal{H}$  denotes the Hamiltonian of the system  $\psi$  the wave function,  $\hbar$  the Planck constant  $h$  divided by  $2\pi$ . The wave function  $\psi$  is a function of the positions of both the nuclei and the electrons, i.e.,  $\psi = \psi(\mathbf{R}, \mathbf{r}, t)$ . Here,  $\mathbf{R}$  denotes positions of the

$k$  nuclei,  $\mathbf{R} = \{\mathbf{R}_1, \dots, \mathbf{R}_k\}$ ,  $\mathbf{r}$  the positions of the  $m$  electrons,  $\mathbf{r} = \{\mathbf{r}_1, \dots, \mathbf{r}_m\}$ , and  $t$  the time.

The idea behind the Born-Oppenheimer approximation is to decouple the fast electron motions from the slow nuclei degrees of freedom (92). The approximation is to assume that the electronic structure adapts instantaneously to given nuclei positions. Within this framework the electrons move in a field of fixed nuclei. The wave function can be separated by the *ansatz*

$$\psi(\mathbf{R}, \mathbf{r}, t) = \psi_n(\mathbf{R}, t)\psi_e(\mathbf{r}; \mathbf{R}), \quad (2.2)$$

with a wave function  $\psi_n(\mathbf{R}, t)$  of the nuclei and the electronic wave function  $\psi_e(\mathbf{r}; \mathbf{R})$  — the latter being only parametrically dependent on the nuclei positions. Given fixed nuclei positions, the electronic wave function may be determined by solving a time-independent Schrödinger equation that contains the nuclei positions  $\mathbf{R}$  only parametrically,

$$\mathcal{H}_e(\mathbf{R}) \psi_e(\mathbf{r}; \mathbf{R}) = E_e(\mathbf{R}) \psi_e(\mathbf{r}; \mathbf{R}). \quad (2.3)$$

Here, the *electronic* Hamiltonian  $\mathcal{H}_e(\mathbf{R}) = \mathcal{H} - \mathcal{T}_n$  with the kinetic energy of the nuclei  $\mathcal{T}_n$  being subtracted from the Hamiltonian of the complete system  $\mathcal{H}$ . Hence,  $\mathcal{H}_e(\mathbf{R})$  contains only derivatives with respect to the positions of the electrons  $\{\mathbf{r}_i\}$ . The eigenvalues of eq. 2.3,  $E_e(\mathbf{R})$ , are termed *potential energy surface* and are the potential that the nuclei “feel” upon motion. Applying eqs. 2.2 and 2.3 into eq. 2.1 yields the the time evolution of the nuclei as a time-dependent Schrödinger equation

$$(\mathcal{T}_n + E_e(\mathbf{R})) \psi_n(\mathbf{R}, t) = i\hbar \frac{\partial \psi_n(\mathbf{R}, t)}{\partial t}. \quad (2.4)$$

As a second step, the translational, rotational, and vibrational degrees of freedom of the nuclei can be separated approximately (93) which is, however, not required for MD simulations. The Born-Oppenheimer approximation holds as long as the eigenvalues of eq. 2.3 significantly differ, i.e., the potential energy surfaces of distinct excited states do not approach each other. For molecules in the ground state this is usually the case.

**(ii) Classical motion of nuclei** Typical macromolecular systems contain up to several hundred thousand atoms which prohibits to solve the time-dependent Schrödinger equation of the nuclei. As a second approximation, MD simulations replace



the nuclei by classical particles that evolve in time according to Newton's second law

$$m_i \frac{\partial^2 \mathbf{R}_i}{\partial t^2} = -\nabla_{\mathbf{R}_i} V(\mathbf{R}) \quad (2.5)$$

$$m_i \mathbf{a}_i = \mathbf{F}_i. \quad (2.6)$$

Here,  $m_i$  denotes the mass of atom  $i$  and  $V(\mathbf{R}) = E_e^{(0)}(\mathbf{R})$  the potential energy surface of the ground state of eq. 2.3. This second approximation can be justified by (a) the fact, that the spacial expectation value of the positions of the nuclei evolve according to the classical equation (Ehrenfest Theorem (94)), and by (b) very short decoherence time scales of proteins in solvent, i.e., a protein in solvent behaves classically (95).

**(iii) Force fields** Due to the large number of electrons in the system, solving the time-independent Schrödinger equation 2.3 is prohibitively expensive. Therefore, the potential energy surface  $E_e(\mathbf{R})$  is approximated by a sum of simple empirical expressions that yield a sufficiently accurate approximation to  $E_e(\mathbf{R})$  but are also computationally cheap to evaluate. The empirical approximation to  $E_e(\mathbf{R})$  is referred to as *force field*. Within the force field, the molecules are approximated by a ball-and-spring-like model. For more details on the current state of force field development we refer to ref. (96).

A typical expression for the potential energy reads (97)

$$V(\mathbf{R}) = V_b + V_a + V_{\text{dih}} + V_{\text{imp. dih}} + V_{\text{LJ}} + V_{\text{coul}} \quad (2.7)$$

$$\begin{aligned} &= \sum_{\text{bonds } i} \frac{k_i}{2} (\ell_i - \ell_{i,0})^2 \\ &+ \sum_{\text{angles } i} \frac{f_i}{2} (\varphi_i - \varphi_{i,0})^2 \\ &+ \sum_{\text{dihedrals } i} \frac{V_i}{2} [1 + \cos(n\phi_i - \phi_{i,0})] \\ &+ \sum_{\text{imp. dih. } i} \kappa_i (\xi_i - \xi_{i,0})^2 \\ &+ \sum_{\text{pairs } i,j} 4\epsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{4\pi\epsilon_0\epsilon_r r_{ij}} \end{aligned} \quad (2.8)$$

Here, the bond stretching potential  $V_b$ , the bond angle potential  $V_a$ , and the improper dihedral potential  $V_{\text{imp.dih}}$  are modeled as harmonic potentials. (Improper dihedrals model bending modes out of a molecular plane such as aromatic rings.) Proper dihedral potentials  $V_{\text{dih}}$  are modeled by a cosine with periodicity  $n$  and potential barriers  $V_i$ . Non-bonded interactions are evaluated pairwise and neglect many-particle effects. Short-range repulsive and attractive dispersion interactions are generally described together by a Lennard-Jones (LJ) potential with the parameters  $\sigma_{ij}$  and  $\epsilon_{ij}$  that determine width and the strength of the potential. In the Coulombic term,  $q_i$  denotes the partial charge of atom  $i$ . The dielectric constant  $\epsilon_r$  is usually set to unity.

The form of the terms in eq. 2.8 differ between different force fields or terms may be replaced by more realistic expressions if required by the application. For example, bond stretching may be modeled by Morse potentials, alkane dihedrals are often modeled by Ryckaert-Bellemans potentials, or the LJ potential can be replaced by the more realistic Buckingham potential (97). For the present thesis, the OPLS all-atom force field was applied (98, 99). Other force fields which are widely used in MD are, for example, different versions of GROMOS (100, 101), AMBER (102), and CHARMM (103, 104). The parameters of the force fields are generally derived by multi-dimensional fits to results from quantum chemical simulations and to thermodynamic quantities such as experimental solvation free energies of amino acids.

Force fields allow to compute interatomic interactions in a sufficiently cheap manner while describing the system in sufficient atomistic detail to investigate, for example, biological macromolecules. However, for any application the validity and accuracy of the force field should be checked, for example by comparison to experimental data. It should be mentioned that biological processes may include chemical reactions, charge transfer processes, or molecules in excited electronic states. Such processes cannot be described by force fields similar equation 2.8, and a more advanced force field than eq 2.8 or even a quantum mechanical description of (parts of) the system may be required.

### 2.1.2. Simulation details

The MD simulations for the present thesis have been performed with the Gromacs simulation package (105, 106). The following paragraphs discuss some details of



MD simulations which are applied to render the simulations both realistic and computationally cheap.

**Time integration** In Gromacs the time integration is performed via a *leap-frog* version of the Verlet algorithm (107) which is mathematically equivalent but computationally more stable than the classical Verlet scheme (108). In the *leap-frog* algorithm, the integration step is computed by

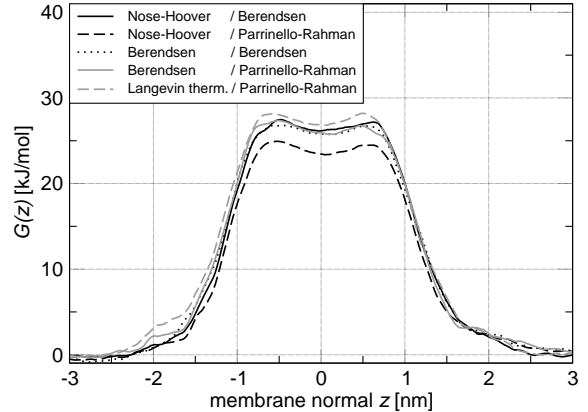
$$\mathbf{v}(t + \Delta t/2) = \mathbf{v}(t - \Delta t/2) + \mathbf{F}(t) \Delta t/m \quad (2.9)$$

$$\mathbf{r}(t + \Delta t) = \mathbf{r}(t) + \mathbf{v}(t + \Delta t/2) \Delta t. \quad (2.10)$$

The time step  $\Delta t$  must be chosen sufficiently small to avoid computational integration errors in the fastest motions, i.e, the vibrations of hydrogen atoms. Explicit simulation of vibrational modes requires a times step of 1 fs. Since bond vibration hardly couple to global motions in proteins, bond lengths are usually constrained during the simulation allowing a time step of 2 fs. For this thesis, the LINCS algorithm (109) was applied to constrain bond lengths of the protein and the lipids, and SETTLE was used for the water molecules (110).

**Lennard-Jones and electrostatic interactions** Pairwise non-bonded interactions scale quadratically with the number of atoms in the system if all possible interactions are explicitly evaluated which is prohibitively expensive for systems of more than  $10^4$  atoms. To avoid the disadvantageous scaling, the LJ interactions are typically cut off between 1.0 and 1.4 nm (103) and a grid-based neighbor searching is applied during the simulation. The most time consuming part in a simulation step is the calculation of the coulomb interactions. Compared to the LJ potential the Coulomb potential decays very slowly with only  $r^{-1}$ . A cut-off in the Coulomb potential can therefore induce severe artifacts (111, 112). The nowadays' standard method to compute electrostatic interactions is via the Particle-mesh Ewald summation which is an improvement to the original Ewald summation. The latter was introduced to calculate long-range interactions of periodic images of crystals (113). Its idea is to split the slowly converging electrostatic energy into long-range and short-range contributions which both converge more rapidly. Since the Ewald summation scales like  $N^2$  with the particle number  $N$ , the method is not practical for large systems. The particle-mesh Ewald (PME) method improves the derivation of

**Figure 2.1:** Potential of mean force (PMF)  $G(z)$  for permeation of a water molecule through a membrane of pure POPC.  $z$  denotes the reaction pathway perpendicular to the membrane. The PMFs have been derived while applying different combinations of temperature/pressure coupling as denoted in the figure legend. The PMFs agree within the statistical uncertainty of 2 kJ/mol.



the long-range electrostatics by assigning the charges to a grid and calculating the potential as a simple sum in reciprocal space(114, 115). Since the switch to reciprocal space can be performed by a fast Fourier transformation the PME method scales like  $N \log N$ .

**Pressure and temperature coupling** Due to the low compressibility of condensed media fixed box dimensions may yield to high pressure during a simulation. Therefore, the simulation box can be coupled to a pressure bath. For the present thesis the Berendsen pressure coupling was applied. The method scales box vectors and coordinates at every step such that the pressure  $P$  decays exponentially towards some reference pressure  $P_0$ ,  $dP/dt = (P_0 - P)/\tau$  (116).

Likewise, the temperature must be controlled during a simulation. Otherwise, numerical inaccuracies, force cut-offs, or dissipative work in non-equilibrium simulations (such as force-probe simulations) can heat up (parts of) the simulation system. Here, the Berendsen algorithm was applied which scales the velocities such that the temperature decays exponentially to the reference temperature (116).

A theoretical drawback of Berendsen pressure and temperature coupling is that they do not generate well-defined ensembles but rather an ensemble in between a canonical and microcanonical ensemble, depending on the coupling parameter  $\tau$  (117). Well-defined ensembles can be generated via Parrinello-Rahman pressure coupling (118, 119), Nosé-Hoover temperature coupling (120, 121), or by coupling of the system to a Langevin thermostat. For practical purposes the difference between the coupling algorithms is small except for observables which do explicitly depend on the velocity distribution or pressure fluctuations. As an example, figure 2.1 shows the potential of mean force (PMF) for water permeation across a





membrane of pure POPC (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine). The PMFs have been derived applying various combinations of temperature/pressure coupling. They agree within the statistical error or 2 kJ/mol (compare sections 2.2 and 4.2).

## 2.2. Potentials of mean force

### 2.2.1. Definition

The concept of potentials of mean force (PMF) was first introduced by Kirkwood in 1935 (122). The PMF is a central quantity in computer studies of biological system since it is directly related to distribution functions and transition rates along reaction pathways. Such pathways may be conformational transitions, chemical reactions, solute permeation, binding or transport events, or any other macromolecular process that can be reduced to one or few reaction coordinates.

The PMF is what its name says: the potential that results from integrating the mean force acting on a particle (or some particles) along some reaction pathway. The present thesis considers permeation events across aquaporin channels. In such cases the reaction coordinate can be chosen as a spatial coordinate of a single particle which helps to keep the notation simple. From the  $3N$  spacial coordinates of the  $N$  particles, let us denote the reaction coordinate as  $x_1$  and all other coordinates as  $x_2, \dots, x_{3N}$ . Then, the mean force acting on the particle being located at  $x_1 = a$  is given by

$$\langle F(x_1 = a) \rangle = \frac{\int dx_2 \cdots dx_{3N} \left( -\frac{\partial V(a, x_2, \dots, x_{3N})}{\partial a} \right) e^{-\beta V(a, x_2, \dots, x_{3N})}}{\int dx_2 \cdots dx_{3N} e^{-\beta V(a, x_2, \dots, x_{3N})}} \quad (2.11)$$

Here,  $V(a, x_2, \dots, x_{3N})$  is the potential energy for the  $N$  particles being located at  $\{a, x_2, \dots, x_{3N}\}$ ,  $-\partial V(a, x_2, \dots, x_{3N})/\partial a$  the force, and  $\exp(-\beta V)$  the Boltzmann factor giving the probability for the system to be in state  $\{a, x_2, \dots, x_{3N}\}$ .  $\beta = 1/k_B T$  is the inverse temperature with the Boltzmann constant  $k_B$  and the temperature  $T$ .

It should be mentioned that for eq. 2.11 we assumed the potential  $V$  to be not explicitly dependent on the momenta of the particles. Otherwise, the integrations

in expression such as eq. 2.11 must additionally be performed over momentum space and the Boltzmann factor would read  $\exp(-\beta H)$  with the the Hamiltonian  $H = T + V$ . If the potential does not depend on the momenta, the contributions from integrating over momentum space cancel between numerator and denominator in eq. 2.11. In the following we assume that this is the case.

The PMF  $W(a)$  along the reaction pathway  $x_1$  is defined naturally from the mean force by

$$\frac{\partial}{\partial a} W(a) = -\langle F(a) \rangle \quad (2.12)$$

### 2.2.2. The PMF is a free energy profile

**Constrained free energies** The following section shows that the PMF *is* a free energy profile. In practice, the two expressions are used synonymously. In statistical physics, the partition function (123) is defined by

$$Z = \int d^{3N}x d^{3N}p e^{-\beta H}. \quad (2.13)$$

The free energy  $\mathcal{W}$  is connected to the partition function via  $\beta\mathcal{W} = \ln Z$ . If the integration in eq. 2.13 is performed over states of constant volume,  $\mathcal{W}$  is referred to as *Helmholtz* free energy, if the integration is performed over states of constant pressure,  $\mathcal{W}$  is denoted as *Gibbs* free energy. A common concept in statistical physics are the so called *constrained free energies* (124). The idea is to describe a statistical system with a prescribed observable  $\mathcal{A}(\mathbf{x}, \mathbf{p})$ . Let us consider a system with some observable  $\mathcal{A}$  being constrained to  $\xi$ . Then, a corresponding partition function is given by

$$Z(\xi) = \int d^{3N}x d^{3N}p \delta(\mathcal{A}(\mathbf{x}, \mathbf{p}) - \xi) e^{-\beta H}. \quad (2.14)$$

where  $\delta$  denotes the Dirac  $\delta$ -function. A free energy of the constrained system may be naturally defined by

$$\beta\mathcal{W}(\xi) = \ln Z(\xi). \quad (2.15)$$



Since the probability for a for every micro state is a priori equal (123), the probability for the unconstrained system to be in a state with  $\mathcal{A} = \xi$  is

$$p(\xi) = \frac{Z(\xi)}{Z} = \frac{1}{Z} e^{-\beta\mathcal{W}(\xi)}. \quad (2.16)$$

**Connection to PMF** To show that  $\mathcal{W}(\xi)$  is the potential of mean force, we consider the case where the constrained observable is  $x_1$ , i.e.  $\mathcal{A}(\mathbf{x}, \mathbf{p}) = x_1$ .  $x_1$  shall be constrained to  $a$ . The derivative of  $\mathcal{W}(a)$  with respect to  $a$  yields

$$\begin{aligned} -\frac{\partial\mathcal{W}(a)}{\partial a} &\stackrel{(2.15)}{=} -\frac{1}{\beta Z(a)} \frac{\partial}{\partial a} Z(a) \\ &= -\frac{1}{\tilde{Z}(a)} \int dx_1 \cdots dx_{3N} e^{-\beta V(x_1, \dots, x_{3N})} \frac{\partial}{\partial a} \delta(x_1 - a) \\ &= -\frac{1}{\beta \tilde{Z}(a)} \int dx_2 \cdots dx_{3N} \frac{\partial}{\partial a} e^{-\beta V(a, x_2, \dots, x_{3N})} \\ &= -\frac{1}{\beta \tilde{Z}(a)} \int dx_2 \cdots dx_{3N} \left( -\beta \frac{\partial V(a, x_2, \dots, x_{3N})}{\partial a} \right) e^{-\beta V(a, x_2, \dots, x_{3N})} \\ &\stackrel{(2.11)}{=} \langle F(a) \rangle. \end{aligned} \quad (2.17)$$

Here, eqs. 2.15 and 2.11 were applied as denoted by the above the equalities. Moreover,  $\tilde{Z} = \int d^{3N}x \exp[-\beta V(\mathbf{x})]$  was introduced to simplify the notation.

Hence, the constrained free energy  $\mathcal{W}(a)$  (defined by eq. 2.15) is the potential of mean force as defined by eq. 2.12. Since the MD simulations for this thesis were performed with pressure and temperature coupling they generated (approximately) *NPT*-ensembles. Therefore, the probability of for the unconstrained system to be in a state with  $\mathcal{A} = \xi$  is given by  $p(\xi) = Z^{-1} \exp[-\beta G(\xi)]$  with Gibbs free energy  $G(\xi)$ . To emphasize in the following that we mean the Gibbs free energy when using the term *free energy*, we use the symbol  $G$  for the latter.

### 2.2.3. Simulation methods to determine the PMF

For macromolecular systems the partition function  $Z$  is usually unknown and the absolute (constrained) free energy cannot be determined. Fortunately, for most applications it is sufficient to determine *relative* change of the free energy along the reaction pathway  $\xi$ . Let  $\xi^*$  denote the reference, then the PMF can be expressed

as

$$G(\xi) = G(\xi^*) - \beta^{-1} \ln \left( \frac{p(\xi)}{p(\xi^*)} \right). \quad (2.18)$$

Here,  $G(\xi^*)$  is an arbitrary constant. In practice,  $G(\xi^*)$  is often set to zero and  $G(\xi)$  is considered as the relative change of the free energy with respect to  $\xi^*$ .  $p(\xi)$  is often referred to as *average distribution function*. According to eqs. 2.14 and 2.16  $p(\xi)$  is given by

$$p(\xi) = \frac{1}{\overline{Z}} \int d^{3N}x \delta(\mathcal{A}(\mathbf{x}) - \xi) e^{-\beta V(\mathbf{x})}. \quad (2.19)$$

In principle, the PMF can be determined from an equilibrium simulation by evaluating  $p(\xi)$  along the reaction path. However, this procedure is often impractical since the sampling at high barriers in the PMF is exponentially small in equilibrium simulations. A possible workaround is to apply artificial forces to either guide (or steer) the system along the reaction pathway, or to confine the system at subsequent positions along the reaction pathway. Subsequently, the artifacts from the guiding or confining potential must be removed to calculate the PMF.

Several methods have been applied to biomolecular system to achieve this task. The following list summarizes the methods. For a more detailed comparison we refer to the concise overview by Kosztin *et al.* (125).

A For **Umbrella sampling simulations** the system is restrained at subsequent positions along the reaction pathway using a restraining potential that remains constant during the simulation (126, 127). The starting positions must be chosen sufficiently close to each other to ensure that the complete reaction pathway is well-sampled. Umbrella sampling was extensively applied for the present thesis and is explained in more detail in the following section.

B In **force-probe MD simulations** (128) (FPMD, or steered MD (129)) the system is restrained to a harmonic guiding potential. During the simulation, the guiding potential moves at constant speed and pulls the system along the reaction pathway. The simulation is repeated several times yielding a distribution  $P(W)$  for the mechanical work  $W$  that was performed by the guiding potential on the system. The challenge is to find a way to reconstruct the underlying PMF from the FPMD simulations. Unfortunately, this task is non-trivial since the work  $W$  is the sum of (i) the free energy difference that



the system has to overcome  $\Delta G$ , and (ii) of dissipative work  $W_d$ ,

$$W = \Delta G + W_d. \quad (2.20)$$

Methods that have been proposed to reconstruct  $\Delta G$  from  $P(W)$  include the following:

- *Jarzynski equality*  $\langle \exp(-\beta W) \rangle = \exp(-\beta \Delta G)$  (130, 131, 132). The validity of the equation depends crucially on a small fraction of sampled pathways with  $W_d < 0$  that transiently violate the *second law* of thermodynamics. Such trajectories are very rare in simulations rendering the method often impractical. The direct application of Jarzynski's equality is therefore only recommended for pulling paths close to equilibrium.
- Jarzynski equality in the stiff-spring approximation (SSA) (130, 133, 134). In the SSA, the free energy difference can be calculated via

$$\Delta G(\xi) = -\beta^{-1} \ln \langle e^{-\beta W(\xi)} \rangle. \quad (2.21)$$

Since an exponentially large number of trajectories are required to evaluate the exponential average, the relation is often approximated by the *cumulant expansion* (133),

$$\Delta G(\xi) = -\beta^{-1} \ln \langle e^{-\beta W(\xi)} \rangle \approx \langle W(\xi) \rangle - \beta \sigma_\xi^2 / 2, \quad (2.22)$$

where  $\sigma_W^2$  is the variance of the work distribution,

$$\sigma_\xi^2 = \langle W(\xi)^2 \rangle - \langle W(\xi) \rangle^2 \quad (2.23)$$

- Forward and reverse pulling with a stiff spring (125). The *transient fluctuation theorem* by Evans and Searles (TFT) (135, 136) (which is a special case of Crooks theorem (137)) relates the work distributions for forward and backward pullings to the dissipative work. In the SSA the work distribution for both the forward and backward pullings are Gaussian (125). Using TFT, the PMF can be approximated by

$$\Delta G(\xi) = (\langle W_F \rangle - \langle W_R \rangle) / 2 \quad (2.24)$$

where  $W_F$  and  $W_R$  denote the work for the forward and backward pulling, respectively.

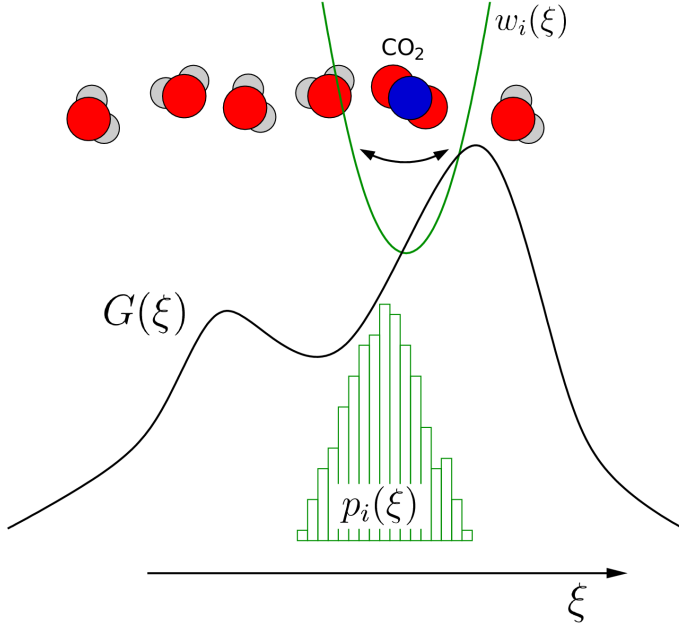
- C For **constrained force simulations** the system is also simulated at discrete positions  $\xi_i$  along the reaction coordinate. In contrast to umbrella sampling, however, the system is reset every step to  $\xi_i$  and the average force  $\langle F(\xi_i) \rangle$  along the reaction coordinate is recorded (138). Subsequently, the average force is integrated along the discrete positions  $\xi_i$ . Hence, constrained force simulations yield the PMF only at discrete positions.
- D Several other methods are frequently used to access free energies such as free energy perturbation (139, 140), thermodynamic integration (141, 142), or particle insertion (143). These methods are particularly useful to compute free energy differences  $\Delta G$  between discrete states, such as a free energy of solvation or ligand binding. To address continuous PMFs along a continuous reaction pathways these methods are less suitable.

#### 2.2.4. Umbrella sampling simulations

A widely used approach to calculate PMFs from computer simulations is the umbrella sampling technique introduced by Torrie and Valleau (126). A comprehensive review can be found in ref. (144).

The key idea is to confine the system at subsequent positions along the reaction coordinate  $\xi$  by applying an additional artificial umbrella potential  $w_i(\xi)$  (compare figure 2.2). The umbrella potential confines the system close to previously selected positions  $\xi_i$  and ensures sufficient sampling around  $\xi_i$ . If the  $\xi_i$  are chosen sufficiently dense, the complete reaction pathway is well sampled when combining the single umbrella runs. The biasing potentials  $w_i(\xi)$  are usually chosen harmonic,  $w_i(\xi) = \frac{1}{2}k(\xi - \xi_i)^2$ , with a force constant  $k$  and centered around successive values of  $\xi_i$ .

Umbrella simulations generate trajectories referring to the potential energy  $V(\mathbf{R}) + w_i(\xi)$ . Hence, the average distribution functions  $p_i(\xi)$  that can be extracted from the simulations are biased by the confining potentials  $w_i(\xi)$ . Using eq. 2.19  $p_i(\xi)$  is



**Figure 2.2:** Illustration of the concept of umbrella sampling simulations. Here, for a  $\text{CO}_2$  molecule permeating along some reaction pathway. To ensure sufficient sampling at barriers along the pathway, the  $\text{CO}_2$  molecule is confined to subsequent positions by applying an artificial harmonic potential.  $G(\xi)$  denotes the unbiased PMF,  $w_i(\xi)$  the umbrella potential,  $p_i(\xi)$  the distribution function (umbrella histogram) extracted from the simulation, and  $\xi$  the reaction coordinate.

given by

$$p_i(\xi) = \frac{\int d^{3N}x \delta(\mathcal{A}(\mathbf{x}) - \xi) e^{-\beta\{V(\mathbf{x})+w_i[\mathcal{A}(\mathbf{x})]\}}}{\int d^{3N}x e^{-\beta[V(\mathbf{x})+w_i[\mathcal{A}(\mathbf{x})]}}. \quad (2.25)$$

Note that  $\mathcal{A}(\mathbf{x})$  is the function that maps the coordinates  $\mathbf{x}$  of the  $3N$  particles onto the reaction coordinate  $\xi$ . Making use of the  $\delta$ -function we get

$$\begin{aligned} p_i(\xi) &= e^{-\beta w_i(\xi)} \frac{\tilde{Z}^{-1} \int d^{3N}x \delta(\mathcal{A}(\mathbf{x}) - \xi) e^{-\beta V(\mathbf{x})}}{\tilde{Z}^{-1} \int d^{3N}x e^{-\beta w_i[\mathcal{A}(\mathbf{x})]} e^{-\beta V(\mathbf{x})}} \\ &= e^{-\beta w_i(\xi)} \frac{p(\xi)}{\langle e^{-\beta w_i[\mathcal{A}(\mathbf{x})]} \rangle} \end{aligned} \quad (2.26)$$

For the last equality we applied eq. 2.19 and the definition of the average  $\langle \dots \rangle = \tilde{Z}^{-1} \int d^{3N}x \dots \exp[-\beta V(\mathbf{x})]$ . Plugging  $p(\xi)$  from eq. 2.26 into eq. 2.18 yields for the unbiased PMF of the  $i$ th umbrella window

$$G_i(\xi) = G(\xi^*) - \beta^{-1} \ln \left( \frac{p_i(\xi)}{p(\xi^*)} \right) - w_i(\xi) + F_i, \quad (2.27)$$

with the undetermined constants  $F_i$  defined via

$$e^{-\beta F_i} = \langle e^{-\beta w_i[A(\mathbf{x})]} \rangle. \quad (2.28)$$

The crucial step is to determine the constants  $F_i$  that correspond to the free energy difference for introducing of the umbrella potentials. Early approaches adjusted the adjacent  $G_i(\xi)$  by constants offsets such they they agree in the regions where they overlap (126, 127). The adjustment can be done, e.g., by a least squares fit in regions with overlap (145). Another approach is to apply free energy perturbation to compute difference of the  $F_i$  between adjacent umbrella windows (146). For the present thesis the *weighted histogram analysis method* (WHAM) (147) was applied which is nowadays considered as the most suitable method to compute the PMF from the umbrella histograms (144).

### 2.2.5. Weighted histogram analysis method (WHAM)

The WHAM procedure was introduced by Kumar *et. al* (147) and is a generalization of the histogram method by Ferrenberg and Swendsen (148). The central idea is to write the (unbiased) distribution function  $p(\xi) \propto \exp(-\beta G(\xi))$  as a weighted sum over the unbiased individual distribution functions  $p_{(i)}^{\text{unbiased}}$ . Subsequently, the weights are chosen to minimize the statistical error of  $p(\xi)$ . Hence, the WHAM procedure selects the overall PMF that (given the umbrella histograms) corresponds to the smallest statistical error and which is therefore considered as the most likely PMF.

**Derivation of the WHAM equations** The following paragraph briefly sketches the derivation of the WHAM equations. The detailed derivation can be found in the original publication (147). From eq. 2.26 the unbiased individual distribution functions can be expressed as

$$p_{(i)}^{\text{unbiased}} = e^{+\beta w_i(\xi)} p_i(\xi) e^{-\beta F_i}. \quad (2.29)$$

The overall distribution function is expressed as a weighted sum over the  $p_{(i)}^{\text{unbiased}}$ ,

$$p(\xi) = \sum_{i=1}^{N_w} \omega_i p_{(i)}^{\text{unbiased}} \quad \text{with} \quad \sum_{i=1}^{N_w} \omega_i = 1. \quad (2.30)$$





Here, the  $\omega_i$  denote the weights and  $N_w$  the number of umbrella windows. The statistical error of  $p(\xi)$  depends on both the weights and the statistical errors of the  $p_{(i)}^{\text{unbiased}}$ , i.e.

$$\delta^2 p(\xi) = \sum_{i=1}^{N_w} \omega_i^2 \delta^2 p_{(i)}^{\text{unbiased}}, \quad (2.31)$$

where the  $\delta^2 p_{(i)}^{\text{unbiased}}$  are estimated from the number of data points of the histogram  $n_i$  and an autocorrelation time which will cancel out in end.  $\delta^2 p(\xi)$  is minimized by setting the derivatives of  $\delta^2 p(\xi)$  with respect to a all  $\omega_i$  equal to zero,

$$\frac{\partial[\delta^2 p(\xi)]}{\partial \omega_i} = 0, \quad i = 1, \dots, N_w. \quad (2.32)$$

Solving the set of equations 2.32, yields the WHAM equations

$$p(\xi) = \frac{\sum_{i=1}^{N_w} n_i p_i(\xi)}{\sum_{i=1}^{N_w} n_i e^{-\beta[w_i(\xi) - F_i]}}, \quad (2.33)$$

where the free energy constants  $F_i$  are calculated from the best estimate of for the distribution function,

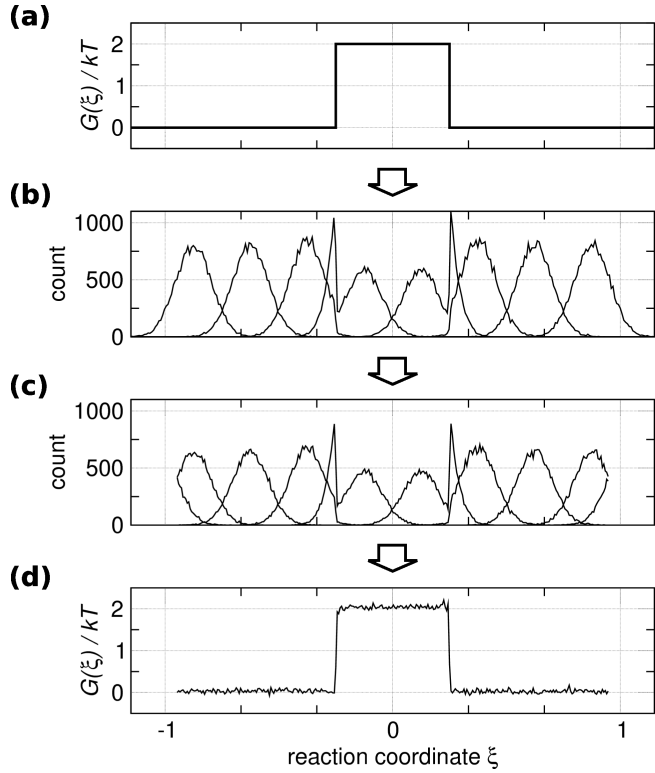
$$e^{-\beta F_i} = \int d\xi e^{-\beta w_i(\xi)} p(\xi). \quad (2.34)$$

The WHAM equations must be solved self-consistently. Depending on the number of histograms  $N_w$ , between  $10^2$  and  $10^5$  iterations are typically required until  $p(\xi)$  has sufficiently converged.

**Cyclic WHAM procedure** Transitions with an inherent periodicity should be described by a periodic PMF. For example, the PMF for the rotation of a dihedral angle exhibits a periodicity of  $2\pi$ . Likewise, the PMF for permeation of a solute across a membrane equals on the two sides of the membrane if the chemical potentials on the two sides of the membrane equal. Due to the periodic boundary conditions this is usually the case in computer simulations of membranes or membrane channels.

To account for the periodicity a cyclic version of WHAM was implemented. To render the procedure cyclic within the interval  $[\xi_{\min}, \xi_{\max}]$ , two modifications

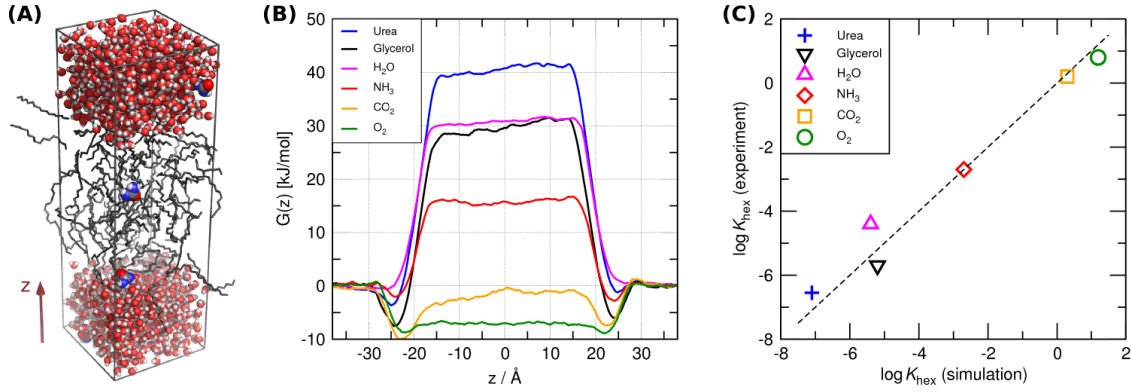
**Figure 2.3:** Validation of the cyclic WHAM procedure through simple step potential. (a) Given step potential  $G(\xi)$ . (b) Simulated umbrella histograms  $p_i(\xi)$ . Each histogram contains 20,000 data points which are distributed according to  $\exp\{-\beta[G(\xi) + w_i(\xi)]\}$ . (c) For demonstration, the periodic boundaries were chosen at  $\xi = \pm 0.95$ . Histograms that expand beyond the boundaries are mapped back into the interval. (d) PMF determined from the histograms in (b) by a cyclic WHAM procedure. The step potential in (a) is well reproduced.



compared to the non-cyclic method are sufficient:

- histograms that extend beyond the boundaries of the interval are mapped into the interval by adding or subtracting  $\Delta\xi = \xi_{\max} - \xi_{\min}$  to the reaction coordinate.
- When evaluating the umbrella potential in eqs. 2.33 and 2.34,  $w_i(\xi) = \frac{1}{2}k(\xi - \xi_i)^2$ ,  $\xi - \xi_i$  is replaced by  $\xi - \xi_i \pm \Delta\xi$  if  $|\xi - \xi_i| > \Delta\xi/2$ .

Figure 2.3 presents the validation of the cyclic WHAM procedure by a simple step potential,  $G(\xi) = 2kT$  for  $\xi < 0.25$  and  $G(\xi) = 0$  otherwise (fig. 2.3a). Umbrella histograms  $p_i(\xi)$  were generated by 20,000 random numbers which were distributed according to  $\exp\{-\beta[G(\xi) + w_i(\xi)]\}$  (fig. 2.3b). Subsequently, a cyclic WHAM procedure was applied on the generated umbrella histograms which reproduces the given step potential (fig. 2.3d).



**Figure 2.4.:** (A) Simulation box of a hexadecane/water interface. For urea molecules which were restrained during the umbrella simulations are shown in sphere representation. (B) Potentials of mean force  $G(z)$  for moving the solute (as denoted in the figure legend) from the bulk across the hexadecane slab. The  $z$ -direction is indicated by the red arrow in figure (A). (C) The hexadecane/water partition coefficient  $K_{hex}$  taken from experiment and calculated from the PMF in (B) via  $K_{hex} = \exp(-\Delta G/k_B T)$ . Here,  $\Delta G$  denotes the free energy difference between the solute in water and in hexadecane.

## 2.3. Validation of solute parameters

For the present thesis extensive umbrella sampling simulations have been performed to calculate PMFs for solute permeation through aquaporin channels or lipid membranes. The solutes that have been simulated include  $H_2O$ ,  $CO_2$ ,  $O_2$ ,  $NH_3$ , glycerol, and urea. During a permeation event of a solute, the solvation shell of the solute needs to be partially or even completely removed from the solute. Hence, the main energetic cost against permeation is due to desolvation which needs to be accurately described by force field parameters of the solute.

Solvation free energies are not available for all solutes. Instead, partition coefficients  $K = c_{apolar}^{sol}/c_w^{sol}$  between water and apolar solvents are mostly available in the literature. Here,  $c_{apolar}^{sol}$  and  $c_w^{sol}$  denote the equilibrium concentrations of the solute in the apolar solvent and water, respectively, if the solvents are in contact.  $K$  is related to the in solvation free energy difference  $\Delta G^{w \rightarrow apolar}$  between the apolar solvent and water via

$$K = \exp(-\Delta G^{w \rightarrow apolar}/k_B T). \quad (2.35)$$

Here, as a preparatory work for chapters 3 and 4, the solute parameters are validated by calculating the hexadecane/water partition coefficients  $K_{hex}$ . The simulation results are compared to experimental values which were taken from references (149, 150, 151, 152). In the simulation,  $K_{hex}$  was calculated by umbrella

simulations at a water/hexadecane interface (figure 2.4A).

The results are summarized in figure 2.4. Figure 2.4B presents the potentials of mean force for pulling a single molecule of  $O_2$ ,  $CO_2$ ,  $NH_3$ ,  $H_2O$ , glycerol, or urea from the water phase across a layer of hexadecane.  $K_{\text{hex}}$  as computed from the PMFs is compared to the experimental  $K_{\text{hex}}$  in fig. 2.4C. The simulations reproduce the experimental  $K_{\text{hex}}$  favorably.

### 3. Gas permeation through Aquaporin-1

Gas transport across biological membranes is of pivotal importance for all living organisms. Physiologically relevant gases are, for example,  $\text{CO}_2$ ,  $\text{O}_2$ ,  $\text{NO}$  or  $\text{NH}_3$ . In plants,  $\text{CO}_2$  exchange is essential for photosynthesis. The human body facilitates the transport of  $\text{CO}_2$  and  $\text{O}_2$  between the pulmonary alveoli and tissues where these gases are required by the metabolism. Along their way,  $\text{CO}_2$  and  $\text{O}_2$  cross a number of membranes including the membranes of the alveolar endothelia and the red blood cell membrane. Nitric oxide  $\text{NO}$  is important for the regulation of blood pressure (153). Moreover,  $\text{NO}$  plays a role in function an disease of the kidney (154), and it acts as a neurotransmitter (155). Ammonia ( $\text{NH}_3$ ) and its conjugated acid  $\text{NH}_4^+$  are the primary substrates for the synthesis of amino acids and are therefore essential for any living cell (31).

At the end of the 19<sup>th</sup> century Overton found that the relative permeability of a lipid bilayer with respect to different solutes is approximately proportional to the solute's oil/water partition coefficient (156). Overton's rule implies that membranes are highly permeable to hydrophobic molecules such as  $\text{CO}_2$ ,  $\text{O}_2$  or  $\text{NO}$ . In spite of its more hydrophilic nature,  $\text{NH}_3$  was also assumed to diffuse rather freely across lipid bilayers.

This view has been challenged since the membranes of several tissues were found to show low  $\text{NH}_3$  permeability, such as membranes in the renal thick ascending limb (157), the colonic crypts (158), or the gastric glands (159, 160). Surprisingly, apical membranes of the gastric glands even have a low permeability to  $\text{CO}_2$  (160) — a finding that seems to contradict Overton's rule. Such unusual permeation characteristics may be necessary to protect the cell from the inhospitable environment of the stomach.

The issue of gas permeation through aquaporins has been a matter debate since it

has been reported that AQP1 enhances the CO<sub>2</sub> permeability of *Xenopus* oocytes (59). In tobacco, the expression of NtAQP1 (a close homolog to human AQP1) was found to enhance CO<sub>2</sub> transport inside the plant and to correlate with leaf growth (70). Other studies doubt a physiological role of CO<sub>2</sub> permeation through AQPs (71, 161). To dissolve the controversy on CO<sub>2</sub> permeation through AQP1 we studied the process using MD simulation as described in detail in the next section. Following the discussion on CO<sub>2</sub> permeation through AQP1, it has more recently been reported that the permeation of O<sub>2</sub> and NO across membranes correlate with inhibition of AQP1 (74, 75). These findings, however, remain to be verified by other research groups. NH<sub>3</sub> has been found to permeate various aquaglyceroporins (60, 61). Likewise to CO<sub>2</sub>, a physiological role of AQP mediated NH<sub>3</sub> transport in mammals has been questioned (162).

## 3.1. Does CO<sub>2</sub> permeate through Aquaporin-1?

### 3.1.1. Motivation

It has been a long standing and controversially discussed question whether CO<sub>2</sub> transport across biological membranes may be facilitated by aquaporins, or whether CO<sub>2</sub> can freely permeate lipid bilayer membranes. The role of AQP1 for CO<sub>2</sub> permeation in mammalian lungs and red blood cells, for example, remains unclear (59, 65, 68, 69, 71, 161, 163, 164). Part of the controversy is apparently due to the fact that (a) CO<sub>2</sub> permeation is usually measured indirectly, via acidification through the action of carbonic anhydrase (CA), (b) that depending on the pH and on the concentration of CA the CO<sub>2</sub> permeation is limited by unstirred water layers on both sides of the membrane (165), (c) that in erythrocytes this acidification is additionally influenced by the action of the HCO<sub>3</sub><sup>-</sup>-Cl<sup>-</sup> transporter, and that (d) inhibitors like 4,4'-diisothiocyanato-stilbene-2,2'-disulfonic acid (DIDS) and mercurial compounds may not be specific for the HCO<sub>3</sub><sup>-</sup>-Cl<sup>-</sup> transporter and aquaporin-1, respectively (65, 69, 161, 163). Moreover, there is a large variability in the CO<sub>2</sub> permeability of different membranes (69, 161, 165, 166).

Another process for which aquaporin-mediated CO<sub>2</sub> permeation has been suggested to play a physiological role is photosynthesis. In a recent study it was shown that the leaf growth of tobacco plants was dependent on the level of NtAQP1 expression, an aquaporin homologous to human AQP1 (70).



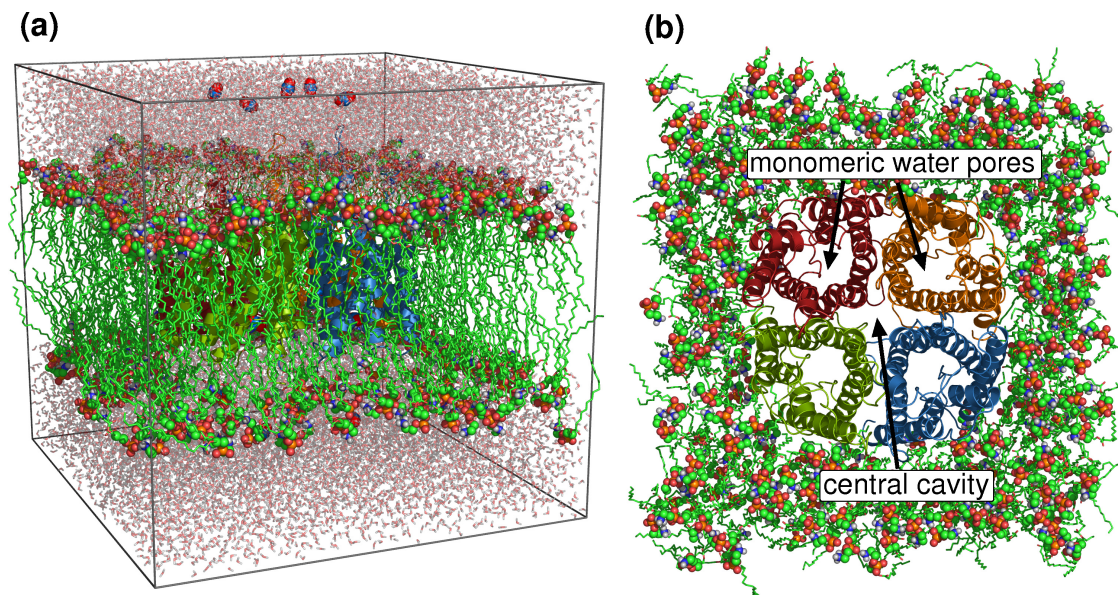
Here, we study the barrier for CO<sub>2</sub> permeation through aquaporin-1 as well as through a pure 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine (POPE) bilayer, using extensive atomistic molecular dynamics simulations. We address the question whether CO<sub>2</sub> is likely to permeate through aquaporin-1 and close homologues like the plant aquaporin NtAQP1 by comparing the free energy for CO<sub>2</sub> permeation through aquaporin-1 to the corresponding profiles for a model lipid bilayer membrane. Moreover, we present present profiles for CO<sub>2</sub> interactions along possible permeation pathways and estimate permeation coefficients.

### 3.1.2. Methods

An equilibrated simulation box of a tetramer of human aquaporin-1 embedded in a solvated lipid bilayer was chosen as starting configuration of the simulations. The starting structure of human aquaporin-1 was modeled based on the x-ray structure of bovine AQP1 (PDB code 1J4N (15)) by mutating differing residues using the WHAT IF modeling software (167). The two sequences are 91% sequence identical, making a structural overlay straightforward. The periodic simulation box contained the aquaporin-1 tetramer, 271 POPE lipids and 19769 TIP4P (168) water molecules. The OPLS all-atom force field (98, 99) was employed for the protein, lipid parameters were taken from Berger *et al.* (169). Four chloride ions were added to neutralize the simulation cell. The system was equilibrated for 2 ns prior to the production simulations. A typical simulation box is shown in fig. 3.1a.

All simulations were carried out using the gromacs simulation software (106, 170). Electrostatic interactions were calculated with the particle-mesh Ewald method (114, 115). Short-range repulsive and attractive dispersion interactions were described together by a Lennard-Jones potential, which was cut off at 1.0 nm. The Settle (110) algorithm was used to constrain bond lengths and angles of water molecules, and lincs (109) was used to constrain all other bond lengths, allowing a time step of 2 fs. The simulation temperature was kept constant by weakly ( $\tau=0.1$  ps) coupling the protein, lipids, and solvent separately to a temperature bath (116) of 300 K. Likewise, the pressure was kept constant by weakly coupling the system to a pressure bath of 1 bar. The  $xy$  (membrane plane) direction was pressure coupled with a coupling constant  $\tau$  of 1 ps, whereas the box length in  $z$  direction was kept fixed to avoid artifacts in the umbrella sampling simulations (see below).

Partial charges for CO<sub>2</sub> were obtained from electrostatic fitting using the CHELPG



**Figure 3.1.:** (a) A typical simulation setup of an AQP1 tetramer, solvated in a POPE bilayer and water. 5 CO<sub>2</sub> molecules in bulk water are shown in blue and red. (b) In the top view, the four monomeric water pores and the central cavity can be identified.

procedure to wavefunctions obtained at the RHF/6-31G\* level, resulting in a charge of  $0.9378e$  on the carbon atom and  $-0.4689e$  on the oxygen atoms. The Lennard-Jones parameters  $\sigma$  and  $\epsilon$  for the CO<sub>2</sub> carbon were 0.375 nm and 0.439 kJ/mol, respectively, and for the CO<sub>2</sub> oxygen 0.296 nm and 0.877 kJ/mol, respectively. The CO<sub>2</sub> parameters were validated by comparison to the oil/water partition coefficient of CO<sub>2</sub>. The experimentally determined value of near unity (171) compares favorably to the small free energy difference for CO<sub>2</sub> between the bulk water phase and the lipid tail region of the bilayer (cf. fig. 3.4, green curve).

**Umbrella simulation setup** The CO<sub>2</sub> simulations were set up as follows: the AQP1 channel was divided into 0.5 Å wide equidistant sections with the center of each section representing an umbrella center. For each section the equilibrated trajectory was searched for snapshots with water molecules as close as possible to the chosen position of the umbrella potential. This water molecule was subsequently replaced by CO<sub>2</sub>. Umbrella sampling calculations were carried out by applying a harmonic restraint force along the pore coordinate with force constants between 100 and 6400 kJ/mol/nm<sup>2</sup>. The umbrella sampling simulations were performed with CO<sub>2</sub> with the restraint acting on the carbon atom, and – for comparison and to





validate the procedure – with water by restraining the oxygen atom. No restraint along the lateral directions was applied.

We observed that a possible flexibility of Arg195 has strong impact on the barrier in the aromatic/arginine (ar/R) constriction region (see results). Since these flexibilities are sampled slowly in the simulations we performed additional umbrella samplings in the ar/R region with starting positions taken from an independent equilibrium run. This way, eight independent channels contributed to our profile. Additionally, the protonation state of His180 opposing the Arg195 might influence the barrier height. In six out of eight channels simulated the  $\delta$ -nitrogen was protonated, in two channels the  $\epsilon^2$ -nitrogen. In accordance with a recent finding for water permeation, no significant influence of the protonation state of His180 on the CO<sub>2</sub> barrier was found (172).

To reduce CPU time, and hence improve sampling, CO<sub>2</sub> molecules were placed in each pore and, to further improve statistics, at three or four positions along the channel coordinate within one simulation run. A minimal distance of 25 Å between CO<sub>2</sub> molecules along the pore was imposed to ensure that interactions between CO<sub>2</sub> molecules (if any) are negligibly small. The umbrella sampling histograms from all monomeric channels were combined to compute a free energy profile using the weighted histogram analysis method (WHAM) (147). In order to account for the periodicity of the system, we implemented a cyclic version of the WHAM procedure. In total, 3912 histograms from 400 ps simulations were obtained (taking only the last 300 ps for analysis and using the first 100 ps for equilibration), extracted from a total of 334 ns of simulation time of the aquaporin system.

To account for fluctuations of each of the channels within the tetramer, the umbrella histogram positions were corrected with respect to the center of the corresponding monomer. As a robust measure for the monomer position we chose the center of mass of the backbone atoms of the transmembrane helices.

Additionally, CO<sub>2</sub> molecules were placed along the central channel surrounding the tetrameric axis (see fig. 3.1a and 3.1b). Umbrella positions were chosen as before, either by replacing a water molecule, or by adding it to the structure in case of an empty position. For the central pore along the tetrameric axis, 888 histograms were obtained.

Umbrella simulations to obtain a free energy profile for water along the AQP1 pore were performed the same way, except that water molecules close to the desired

minimum of the artificial potential were not replaced but were directly restrained. Like for CO<sub>2</sub>, the individual umbrella simulations were carried out for 400 ps. In total, 3092 histograms were obtained. Force constants were chosen between 100 and 1600 kJ/mol/nm<sup>2</sup>.

Umbrella simulations for the lipid bilayer were performed by inserting CO<sub>2</sub> molecules at various positions into random snapshots of a 20 ns equilibration run of a bilayer patch of 128 POPE and 4777 TIP4P water molecules. Up to 12 CO<sub>2</sub> molecules were inserted into one simulation keeping a minimum distance of 25 Å in the lateral direction and 30 Å perpendicular to the bilayer. A total of 3200 histograms were extracted from 260 simulations of 500 ps each, using the last 300 ps for analysis and the first 200 ps for equilibration. Force constants between 100 and 800 kJ/mol/nm<sup>2</sup> were used.

**Enthalpic interactions** Profiles for enthalpic interactions of CO<sub>2</sub> (fig. 3.5) were calculated as the sum of short-range Coulomb interactions ( $\leq 1$  nm) and Lennard-Jones interactions. Note that these do not correspond to the total enthalpy for a given CO<sub>2</sub> position since they only include interactions directly involving CO<sub>2</sub>. CO<sub>2</sub> induced alterations in the surroundings (like protein-water interactions) are not included in this component of the enthalpy. The profiles do, however, give more insight into the origin of the free energy barriers.

**Permeability estimations** Permeability coefficients  $P_f$  for the POPE membrane were estimated by counting barrier crossings in simulations where CO<sub>2</sub> molecules were allowed to diffuse freely through a POPE bilayer (see below). After equilibration, we observed 3.4 barrier crossings per barrier, direction and nanosecond. The permeability coefficient for *one* barrier of  $\Delta G^{POPE} = 4$  kJ/mol was calculated by  $P_f^{(1)} = \Phi / (S c_s)$  where  $\Phi = 3.4/\text{ns}$  denotes the transition rate,  $S = 30 \text{ nm}^2$  our membrane surface and  $c_s = 0.52 \text{ M}$  the CO<sub>2</sub> concentration, giving  $P_f^{(1)} = 23 \text{ cm/s}$ . Since half of the CO<sub>2</sub> molecules crossing the first barrier will return and not permeate through the bilayer this refers to an estimated permeability coefficient for POPE of  $P_f^{POPE} = 12 \text{ cm/s}$ .

Using the permeability coefficient for a single POPE barrier we estimated the coefficients for the water pore and the central cavity via  $P_f' = P_f^{(1)} \exp(-(\Delta G' - \Delta G^{POPE})/kT)$ . Here,  $\Delta G'$  denotes the barrier of the water pore or the central cavity,  $T$  the temperature and  $k$  the Boltzmann constant. In terms of rate theory,



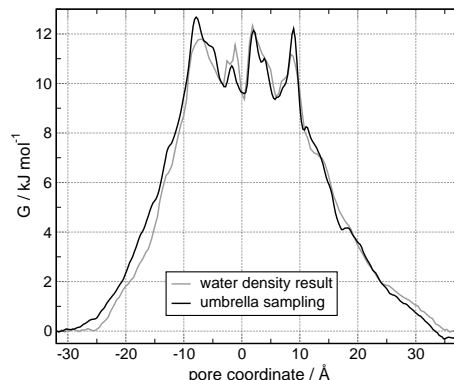
this estimate assumes the same prefactor for CO<sub>2</sub> permeation across a single POPE barrier as through a channel. Additionally, it rests on the assumption that the permeation rate is limited solely by the main barrier along the reactive coordinate. Along the central cavity two barriers of the same height are present, giving rise to an additional factor of 1/2. Unit permeability coefficients were calculated by  $p_f = P_f S_{\text{mono}}$  since the free energy barriers were normalized to an area of one AQP monomer. We estimated the area of one monomer as  $S_{\text{mono}} = 10.3 \text{ nm}^2$ . Note that the area per monomer *in vivo* can be expected to be lower than the area per monomer in AQP1 crystals ( $11.5 \text{ nm}^2$ ) (28).

**Error analysis** Statistical errors of the free energy profiles were estimated using bootstrap analysis. In order to account properly for limited long time sampling we performed bootstrap analysis by considering complete histograms as independent data points: From the  $N$  original histograms we randomly selected  $N$  histograms (allowing duplication) which were subsequently used for WHAM analysis. This procedure was performed 50 times for each profile allowing the calculation of standard deviations. The standard deviations at the main barriers were found to be approx. 1 kJ/mol for the monomeric channel, 2 kJ/mol for the central cavity, and 1 kJ/mol for the POPE membrane.

The barrier at the ar/R region in the monomeric channel is influenced strongly by the flexibility of Arg195 which is sampled slowly (see above). Therefore, we calculated the barrier for each of the eight simulated channels separately. Considering these eight barriers as independent data points gave an additional uncertainty of 3 kJ/mol.

**Bulk water correction** Umbrella simulations such as presented here, with a molecule restrained to a certain  $z$  position, i.e., to an  $x$ - $y$ -plane parallel to the bilayer, will not sample all possible  $x$ - $y$  positions in the simulated time window, especially within the bulk water region. This leads to an overestimation of the free energy level in the bulk region. The correction for this entropic effect is straightforward, as the effect is related to the sampled area (and hence, the simulation time). We therefore took advantage of the fact that in AQP1 the relative free energy for water can be obtained not only by umbrella simulations but also from the water number density distribution via  $G(z) = -kT \ln \rho(z)$ . The bulk free energy level for water can be easily and accurately determined from such a profile. A 14 ns equilibrium

**Figure 3.2:** Free energy profiles of water permeating through the monomeric AQP1 pore obtained from umbrella simulations (black) and from the evaluation of the water density distribution (shaded).



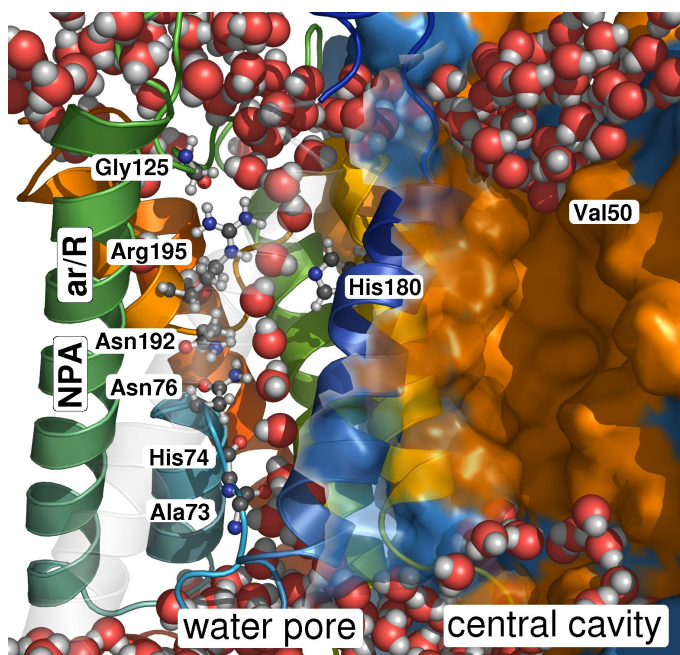
MD simulation was used for this purpose.

The bulk level was normalized to the area of an AQP1 monomer, estimated as approximately  $A_{\text{mono}} = 10.3 \text{ nm}^2$ , to allow direct comparison to the lipid membrane profile, which is independent of the area. The two profiles derived from the umbrella sampling simulations and from the water density distribution were found to be rather similar (fig. 3.2), rendering an overlay of the two profiles straightforward. The resulting bulk level correction for the umbrella sampling simulations of water of  $\Delta G_{\text{w}}^{\text{bulk}} = 3.2 \text{ kJ/mol}$  was found to be relatively small. The correction is already incorporated in Fig. 3.2 as a linear trapezoidal correction to the umbrella profile in the pore entrance and exit region.

The bulk correction depends on the sampled area within the simulation which is approximately proportional to the diffusion constant of the solute. From simulations of a box of pure TIP4P water and of 20  $\text{CO}_2$  molecules solvated in 4084 TIP4P water molecules we calculated the diffusion constants of water and  $\text{CO}_2$  in the simulations to  $D_{\text{w}} = 3.6 \times 10^{-5} \text{ cm}^2/\text{s}$  and  $D_{\text{CO}_2} = 2.7 \times 10^{-5} \text{ cm}^2/\text{s}$ , respectively, giving a bulk correction for  $\text{CO}_2$  of  $\Delta G_{\text{CO}_2}^{\text{bulk}} = \Delta G_{\text{w}}^{\text{bulk}} + kT \ln(D_{\text{w}}/D_{\text{CO}_2}) = 3.9 \text{ kJ/mol}$ .

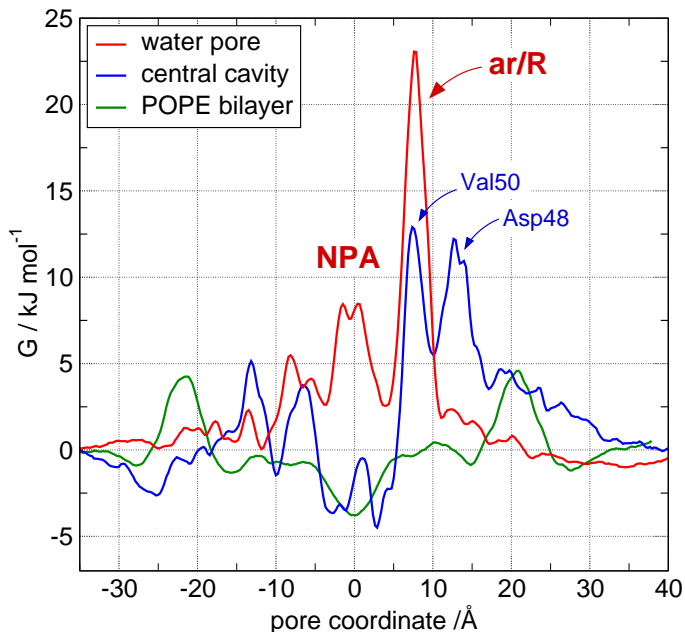
### 3.1.3. Results and Discussion

Figure 3.4 shows the free energy profile obtained as a potential of mean force from a set of umbrella sampling simulations for three potential pathways for  $\text{CO}_2$  permeation through aquaporin-1 embedded in a lipid bilayer: the pathway through an aquaporin water pore (black, solid line), the central channel (black, dashed line), and through the POPE membrane (grey). In the monomeric water channel  $\text{CO}_2$  molecules encounter a major barrier of approx.  $23 \pm 4 \text{ kJ/mol}$  in the aromatic/arginine (ar/R) constriction region (also termed selectivity filter, see fig. 3.3)



**Figure 3.3.:** A snapshot taken from an equilibrium simulation of aquaporin-1 showing the water pore (left, helices and ribbon representation) filled by a single file of water. Several water-interacting residues are displayed in ball-and-stick representation. The aromatic/arginine (ar/R) constriction region forms the narrowest part of the pore. The two highly conserved Asn-Pro-Ala (NPA) motifs are located at the center of the channel. The central cavity along the fourfold axis of the tetramer is shown in surface representation (right). The surface representation is colored according to residue hydrophobicity: hydrophobic residues in orange, hydrophilic residues in blue.

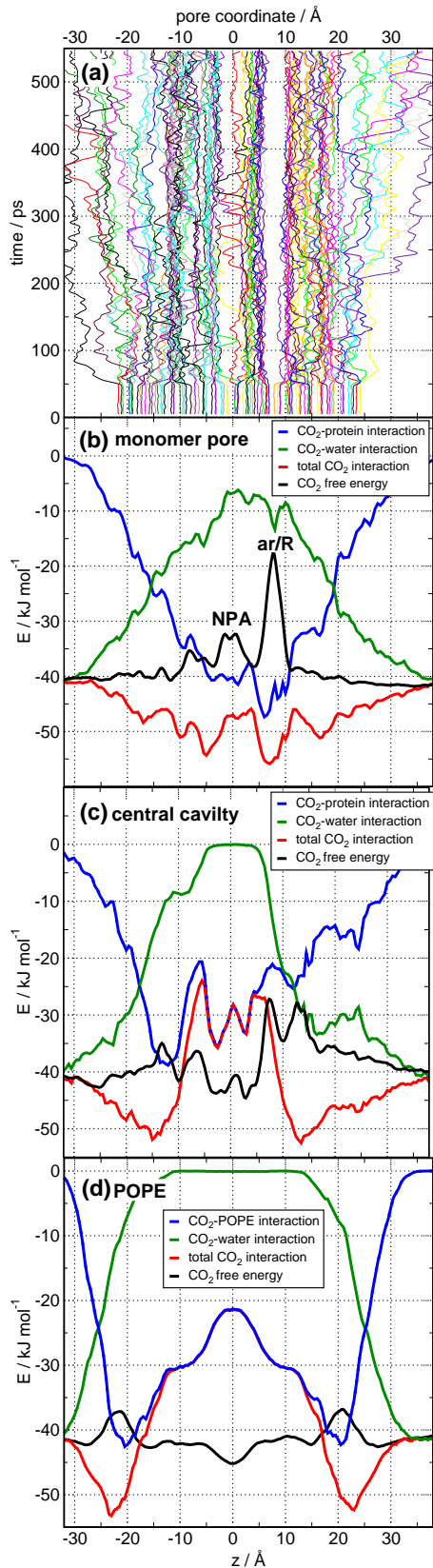
and a lower one of about  $9 \pm 1$  kJ/mol around the region with the two conserved NPA motifs. The corresponding enthalpic interactions of  $\text{CO}_2$  to water and to the protein, respectively, are shown in figure 3.5b. As can be seen, there is no significant loss of interactions involving  $\text{CO}_2$  in the ar/R region, rendering the free energy barrier either entropic in nature, or, more likely, due to indirect enthalpic effects like an unfavorable configuration of surrounding water molecules. This is illustrated in fig. 3.6 which shows a typical pathway of  $\text{CO}_2$  through the monomeric AQP1 channel, as derived from the umbrella sampling simulations. As can be seen, water molecules form strong hydrogen bonds to Arg195 in the ar/R region, which are broken upon  $\text{CO}_2$  passage. Fig. 3.5a shows 115 pathways for  $\text{CO}_2$  molecules that were not restrained by an umbrella potential. The trajectories show that the  $\text{CO}_2$  molecules behave as expected from the potential of mean force, not spontaneously passing the ar/R constriction region within the simulated time of 500 ps.



**Figure 3.4:** Free energy profile for  $\text{CO}_2$  permeation through the aquaporin-1 water pore (red), the tetrameric central cavity (blue) and a POPE bilayer (green).

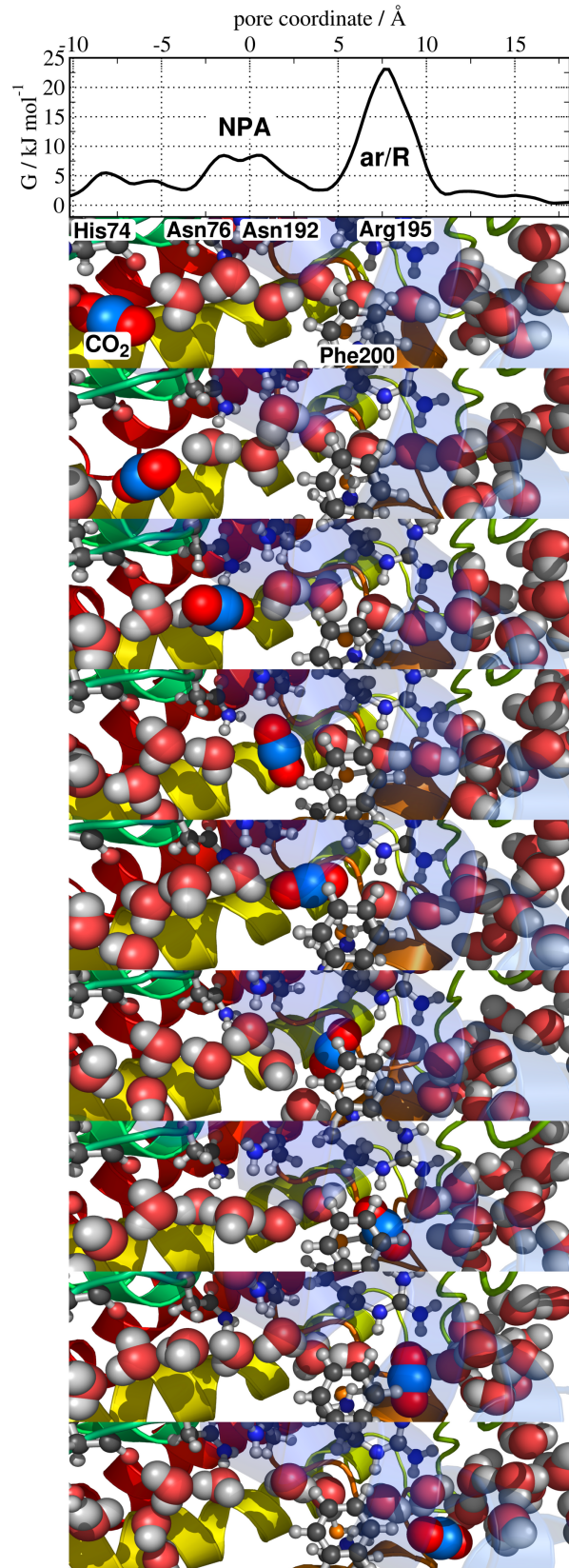
The barrier in the ar/R is strongly influenced by the flexibility of Arg195. In the crystal structure, its guanidinium group points straight into the pore creating the narrowest part of the channel. Two water molecules forming H-bonds to the  $\epsilon$ - and  $\eta^1$ -nitrogens are displaced by a permeating  $\text{CO}_2$  molecule (compare fig. 3.6, 2<sup>nd</sup> and 3<sup>rd</sup> picture from bottom). In this configuration, the barrier is likely to be higher than 23 kJ/mol. However, during simulation we observed some flexibility of Arg195 probably caused by flexibility in Loop C. Arg195 is stabilized sterically by the neighboring Asn127 and by a frequent H-bond to the carbonyl oxygen of Gly125 (see fig. 3.3). In such cases, no breakage of H-bonds between water and Arg195 is necessary lowering the barrier for  $\text{CO}_2$  passage significantly. The effect of the flexibility of Arg195 on the  $\text{CO}_2$  barrier is reflected by the reported uncertainty of 4 kJ/mol.

As described in the methods section we roughly estimated the unit permeability coefficient for  $\text{CO}_2$  permeation through AQP1  $p_f = 1 \times 10^{-15} \text{ cm}^3/\text{s}$ . The error of 4 kJ/mol in the main barrier corresponds to factor of 5 in the permeabilities, i.e. a range of  $2 \times 10^{-16}$  to  $5 \times 10^{-15}$ . The unit permeability coefficient for water permeation through a 2D AQP1 crystal is  $p_f^{\text{water}} = 5.43 \times 10^{-14} \text{ cm}^3/\text{s}$  (28), thus, the resistance of AQP1 to  $\text{CO}_2$  is between one and two orders of magnitude higher in comparison to water. This result can also be expected by comparison of the corresponding free energy barriers of 12 and 23 kJ/mol, respectively.



**Figure 3.5:** Trajectories of unrestrained CO<sub>2</sub> molecules in the monomeric AQP1 channel (a), and profiles of enthalpic interactions of CO<sub>2</sub> with water (green), protein or POPE, respectively (blue), and total (red) together with the free energy profiles of (cf. Fig. 3.4, black) for the monomeric channel (b), the central channel around the tetramer axis (c), and through the POPE bilayer (d). The enthalpic profiles only show interactions involving CO<sub>2</sub> and do not contain interactions within the CO<sub>2</sub>'s surroundings.





**Figure 3.6:** A  $\text{CO}_2$  at various positions along the water channel indicating a possible pathway for  $\text{CO}_2$  along the NPA motifs and through the aromatic/arginine constriction region. The  $\text{CO}_2$  molecule is colored in red and marine blue. On top of the sequence the corresponding free energy is plotted, indicating the 23 kJ/mol barrier putatively caused by competition for hydrogen bonds with Arg195. Note that frequent hydrogen bonds between water molecules and Arg195 break upon  $\text{CO}_2$  passage.





The single channel permeability estimated from our results can be compared with the results from Nakhoul *et al.* (59), Prasad *et al.* (164), and Yang *et al.* (71) from experiments with oocytes, erythrocytes, and proteoliposomes containing AQP1. As discussed by Yang *et al.* (71), the results reported by Nakhoul *et al.* (59) refer to a single channel CO<sub>2</sub> permeability of  $1\text{-}2 \times 10^{-14}$  cm<sup>3</sup>/s whereas the results reported by Prasad *et al.* (164) give  $\sim 5 \times 10^{-15}$  cm<sup>3</sup>/s. Yang *et al.* (71) did not observe AQP1 mediated CO<sub>2</sub> permeation and reported an upper limit of  $3 \times 10^{-16}$  cm<sup>3</sup>/s. Therefore, our estimate of  $\sim 1 \times 10^{-15}$  cm<sup>3</sup>/s most closely matches the result by Prasad *et al.* and the upper limit by Yang *et al.*

A second possible pathway through AQP1 would be the central pore along the fourfold axis of the tetramer. Due to its hydrophobicity the central pore is neither filled by water nor does it conduct water (see fig. 3.3). Two major barriers for CO<sub>2</sub> permeation can be identified: First, a barrier of about  $12 \pm 2$  kJ/mol at  $z = 7.5\text{\AA}$  is located near the extracellular entrance to the central cavity, surrounded by the 4 Val50 of the 4 monomers. In this region CO<sub>2</sub> molecules lose favorable interactions to neighboring water molecules (compare figure 3.5c, green and red lines). A second barrier of the same height is found around  $7\text{\AA}$  above the first one. Here, the CO<sub>2</sub> is surrounded by the 4 Asp48, and hence, it is likely that CO<sub>2</sub> competes with water for hydrogen bonds to Asp48. The barrier translates into a permeability coefficient of  $P_f = 0.1$  cm/s for the central cavity, referring to the area of an AQP1 tetramer, or a unit permeability coefficient of  $p_f = 4 \times 10^{-14}$  cm<sup>3</sup>/s.

The free energy barrier for CO<sub>2</sub> permeation through the central cavity is significantly smaller than for the monomeric channel. When taking into account that each tetramer contributes four monomeric channels and one central pore, these barriers would imply that approx. 10% of AQP1-permeated CO<sub>2</sub> would be expected to permeate via the monomeric channels. Note, however, that in these calculations we have assumed an empty central channel. Should an ion or organic molecule be bound in the central cavity under physiological conditions (like in the recently resolved structure of an aquaporin-Z tetramer (173)), this would be expected to drastically increase the barrier for CO<sub>2</sub> permeation through the central pore. Experiments with mercurial AQP1 blockers and with the C189S mutant suggest that AQP1-mediated CO<sub>2</sub> permeation takes place through the monomeric pore (68, 164), which would imply that, indeed, the central channel is blocked under the experimental conditions.

The tobacco aquaporin NtAQP1 was recently reported to facilitate physiologically relevant CO<sub>2</sub> permeation (70). The pore region of NtAQP1 is highly similar to human AQP1, with the residues surrounding the major barriers in the NPA and ar/R region fully conserved among the two proteins (70). Our results suggest that – if the central channel is blocked – aquaporin-mediated CO<sub>2</sub> permeation is expected to play a significant role in photosynthesis only if the plant cell membranes have a low intrinsic CO<sub>2</sub> permeability, with an activation barrier well above 20 kJ/mol. However, since the structure of the central pore of NtAQP1 is unknown, we cannot exclude that permeation via the central pore lowers the effective barrier.

### Membrane permeability

In order to address the physiological significance of CO<sub>2</sub> permeation through aquaporin-1, we calculated the free energy barrier for CO<sub>2</sub> permeation through a lipid bilayer membrane composed of pure POPE (Fig. 3.5d). Two barriers, one per leaflet, of about 4 kJ/mol were observed at the intersection between the polar headgroups and the aliphatic chains of the lipids, indicating that the POPE membrane is highly permeable to CO<sub>2</sub>. As can be seen from Fig. 3.5d, the free energy barrier coincides with an enthalpy minimum, rendering the free energy barrier either entropic or due to indirect, water mediated enthalpic effects. A set of simulations in which CO<sub>2</sub> was allowed to diffuse freely corroborated the small barrier for CO<sub>2</sub> permeation across the POPE membrane. In 8 simulations of 11 ns each, 112 barrier crossing were observed in total for the 92 CO<sub>2</sub> molecules that were simulated. Inside the membrane, the CO<sub>2</sub> molecules diffused rapidly, i.e. the energetic barriers near the lipid headgroups limit the permeation through the bilayer. The permeability constant calculated from these simulations is  $P_f^{POPE} = 12$  cm/s.

This value is about eight times higher than the CO<sub>2</sub> permeability for artificial membranes reported by Prasad *et al.* (164). The discrepancy could be due to (a) limitations of the force field used in the simulation, (b) the different lipid composition and (c) remaining unstirred layer effects in experiments even at high concentrations of HCO<sub>3</sub><sup>-</sup> and carbonic anhydrase (CA). Given a membrane with intrinsic permeability  $P_f^M$ , surrounded by two unstirred layers of thickness  $d$  each, the total permeability  $P_f^{\text{tot}}$  obeys  $(P_f^{\text{tot}})^{-1} = (P_f^M)^{-1} + (D/2d)^{-1}$ .  $D$  denotes the diffusion constant of CO<sub>2</sub> in water,  $D/d$  can be considered as a permeability of a single unstirred layer. Outside the unstirred layers the CO<sub>2</sub> concentration is con-



sidered constant due to some buffer. Using this relation, a diffusion constant of  $2 \times 10^{-5} \text{ cm}^2/\text{s}$ , two unstirred layers of thickness 60 nm each and an intrinsic membrane permeability of 12 cm/s yield a total permeability of approx. 1.5 cm/s, i.e. approximately the value reported by Prasad *et al.* for artificial membranes. The CA concentration of 0.5 mg/ml used by Prasad *et al.* refers to one CA enzyme per  $(45 \text{ nm})^3$ . Therefore, if CA is not bound to the membrane, the assumption that  $\text{CO}_2$  needs to diffuse through a water layer of 60 nm before entering the membrane is within a realistic order of magnitude. This analysis implies that in experiments,  $\text{CO}_2$  permeation through membranes with *high* intrinsic permeability like POPE might be limited by the diffusion of  $\text{CO}_2$  from CA to the membrane and from the membrane to CA inside the liposome, even at high pH (or high  $\text{HCO}_3^-$  concentration) and high CA concentration. Therefore, intrinsic membrane permeabilities might be underestimated by experiments under such conditions.

### 3.1.4. Conclusions

Taken together, these results suggest that  $\text{CO}_2$  permeation through AQP1 can be expected to play a physiological role only in membranes with a low intrinsic  $\text{CO}_2$  permeability, such as membranes of cells exposed to an inhospitable environment (69, 166). Membranes with similar physicochemical characteristics to POPE are highly permeable to  $\text{CO}_2$ , rendering a physiological role for AQP1-mediated  $\text{CO}_2$  permeation in such membranes unlikely. The role of aquaporin-1 mediated  $\text{CO}_2$  permeation in red blood cells, mammalian lungs and other tissues therefore depends on the intrinsic  $\text{CO}_2$  permeability of the surrounding membrane. Only for membranes with an activation barrier well above 20 kJ/mol or an empty central channel can AQP1-mediated  $\text{CO}_2$  permeation be expected to play a major role, if a high enough expression level is provided. An interesting test of this hypothesis would be an evaluation of the  $\text{CO}_2$  permeability of vesicles with different lipid composition and different levels of embedded AQP1.

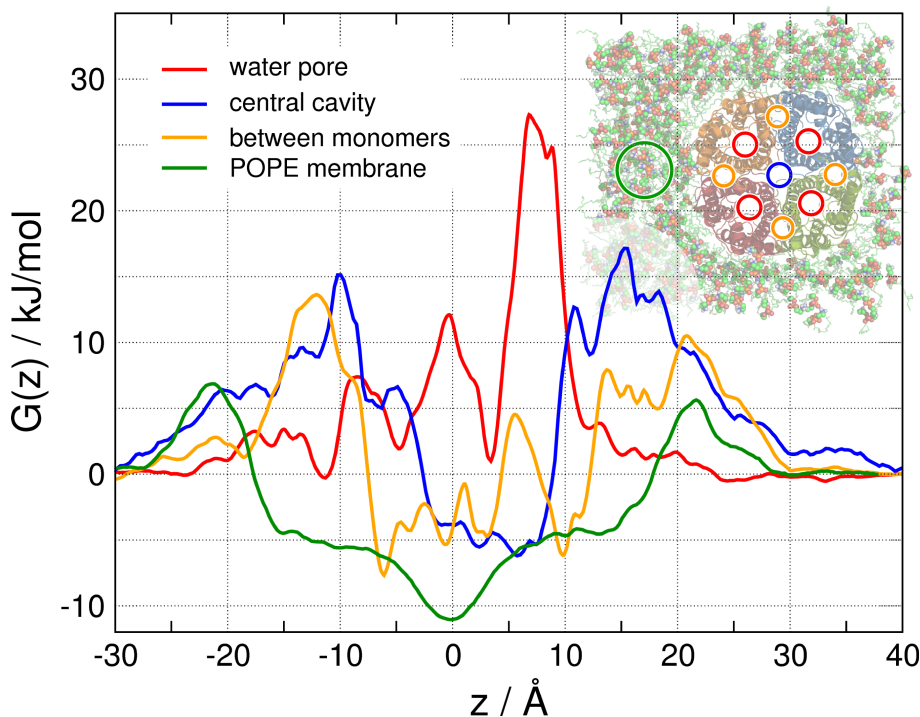
## 3.2. $\text{O}_2$ permeation through Aquaporin-1

Likewise to  $\text{CO}_2$ , molecular oxygen permeates across the membranes of the pulmonary alveoli and the red blood cells. These membranes are known to exhibit a high density of AQP1 which lead to speculations that  $\text{O}_2$  permeation across the

membranes may be facilitated (in part) by AQP1 (75). The speculations have been supported by experiments on lipid monolayers which suggested that  $O_2$  permeability of membranes might be lower than previously thought (174). However, several studies using spin labeling techniques reported very high  $O_2$  permeability for biological membranes including thylakoid membranes of spinach chloroplasts (175), bilayers made of calf lens lipids (176), and bilayers composed of POPE/Cholesterol mixtures (177). Such experiments render a physiological role of AQP1-mediated  $O_2$  permeation unlikely.

To resolve the contradiction we have computed PMFs for  $O_2$  permeating through a tetramer of hAQP1 embedded in a lipid bilayer of pure POPE. The PMFs have been derived as described in section 4.2 and are being compared in fig. 3.7. The PMFs have been calculated for four possible pathways across the bilayer:

- (i) along AQP1 water pore (red),



**Figure 3.7.:** Free energy profiles for four putative pathways for  $O_2$  permeating through an aquaporin-1 tetramer embedded in a POPE bilayer: for the AQP water pore (red), the central cavity (blue), a cavity located between the AQP monomers, and for the pathway across the lipid bilayer (green). The inset shows the AQP1 tetramer embedded in the POPE membrane. The pathways are indicated by circles which are colored according to the curves.



- (ii) the tetrameric central cavity (blue),
- (iii) a cavity located between the AQP monomers (orange) which has recently been proposed as a possible O<sub>2</sub> pathway (178), and
- (iv) through the POPE bilayer (green).

In the inset of fig. 3.7 the four pathways are indicated by circles which are colored according to the curve in the figure.

The maximum barriers along the central cavity (17 kJ/mol) and along the cavity between the monomers (13.5 kJ/mol) are substantially lower than the barrier in the ar/R region of the water pore (27.5 kJ/mol). The POPE bilayer with a maximum barrier of only ~6 kJ/mol is highly permeable to O<sub>2</sub> in agreement with references (175, 176, 177). Hence, none of the hAQP1 cavities contributes significantly to O<sub>2</sub> flux if hAQP1 is embedded in a membrane with similar characteristics to POPE. Only when hAQP1 is embedded in a membrane with low intrinsic permeability for small apolar molecules, and given that the central cavity or the cavity between the monomers are not blocked by an ion or an organic molecule, the cavities could contribute to O<sub>2</sub> permeation across the membrane. From the present data however, a role of hAQP1 as a physiological O<sub>2</sub> channel seems unlikely.

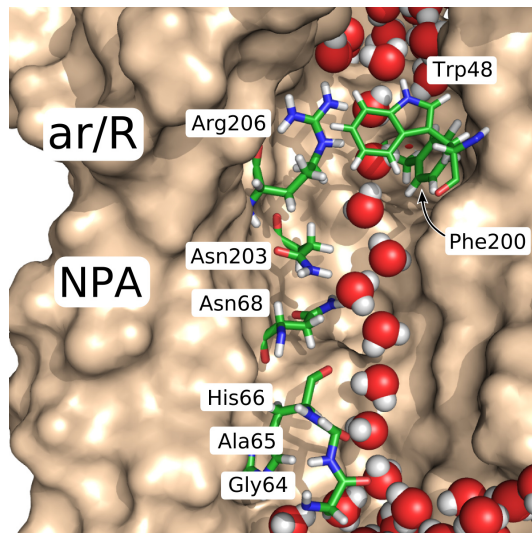


# 4. The selectivity mechanism of aquaporins and aquaglyceroporins

## 4.1. Introduction

The elucidation of the structure of aquaporins (15, 179) and aquaglyceroporins (13) gave the first insights into their selectivity mechanisms and formed the basis for the analysis of dynamics and energetics of water and glycerol conduction (30, 38, 134). Numerous experiments revealed diverse permeation characteristics of aquaporins. Some aquaglyceroporins are permeable to both glycerol and urea or even arsenite (6). Others such as the glycerol facilitator GlpF are permeable for various polyols but hardly for urea (79). The permeability with respect to water differs significantly within the aquaporin family, AQP6 can act as an anion channel (8), and some AQPs were reported to be permeated by ammonia (60). CO<sub>2</sub> permeation through AQPs has been a matter of debate since AQP1 was observed to enhance CO<sub>2</sub> flux (59). Recently, permeation of O<sub>2</sub> and NO across membranes has been reported to correlate with inhibition of AQP1 (74, 75).

Two main constriction sites have been identified in the aquaporin channels (13, 15, 179). In the center of the pore, two conserved Asn-Pro-Ala motifs (NPA, compare fig. 4.1) with their two asparagine side chains pointing into the pore are located at the end of two half helices. The dipoles of the half helices generate an electrostatic barrier in the NPA region, which is, together with desolvation effects, essential for proton exclusion (41, 43, 46, 180). Close to the extracellular exit of the channel, the aromatic/arginine (ar/R) constriction region forms the narrowest part of the pore and is therefore generally assumed to be important for the channel selectivity, an interpretation that has been supported by theoretical studies (181, 182) and mutation experiments (183). The residues in the ar/R region differ in the aquaporin family rendering the constriction site diverse in size

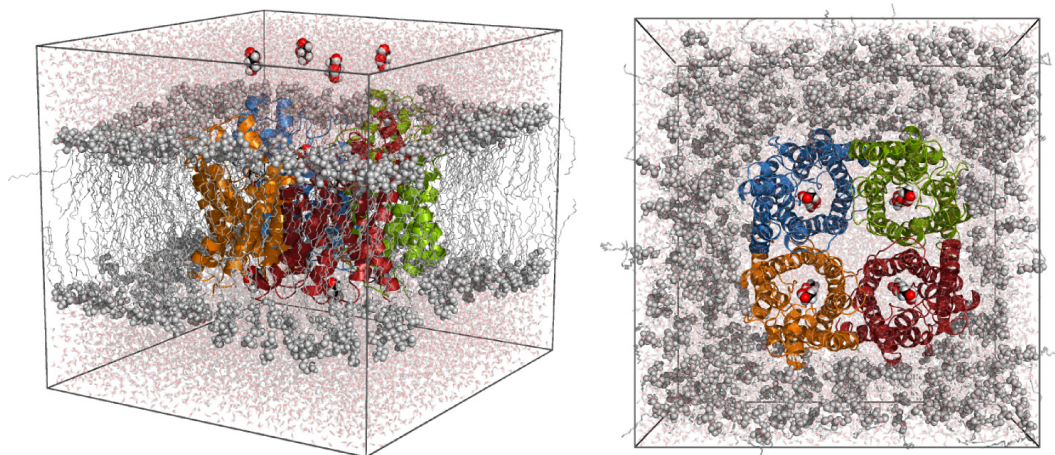


**Figure 4.1:** A snapshot of an MD simulation showing a single file of water inside the pore of the bacterial glycerol facilitator GlpF. Some water-interacting residues are shown in stick representation.

and hydrophobicity. The highly conserved arginine together with the histidine in water-specific aquaporins were considered to be essential for the isolation of water molecules from their solvation shell in the bulk (15). This interpretation has been questioned by recent experiments on oocyte permeability (183).

However, a unifying picture that explains the diverse permeation properties of both aquaporins and aquaglyceroporins has not evolved. Therefore, we applied molecular dynamics simulations to determine permeation barriers of a wide range of solutes permeating through human AQP1 (hAQP1) and GlpF as typical members of the two AQP subfamilies. Potentials of mean force (PMFs) for  $O_2$ ,  $CO_2$ ,  $NH_3$ , glycerol, urea, and water were derived using the technique of umbrella sampling simulations. Since these solutes differ strongly in hydrophobicity and size, the permeation barriers allow to derive a unifying picture of the selectivity of aquaporins and aquaglyceroporins. The physiological relevance of permeation through aquaporins is investigated by comparing the pathway through the aquaporin pores with the pathway across two distinct model membranes. This approach also allows to rationalize permeation experiments on aquaporins and aquaglyceroporins embedded in oocyte membranes or liposomes. The direct evaluation of interactions that govern the selectivity allows to identify the molecular determinants of selectivity. The results reveal a comprehensive mechanism underlying the permeation characteristics of aquaporins and aquaglyceroporins.





**Figure 4.2.:** Simulation box of a GlpF tetramer solvated in a membrane of POPE and water in side view (left) and top view (right). The membrane is shown as grey sticks and spheres, water as transparent red and white sticks. Glycerol molecules (displayed in sphere representation) are placed along the channel axes as starting configurations for umbrella sampling simulations.

## 4.2. Methods

### 4.2.1. Simulation setup

The simulation boxes of hAQP1 or GlpF contained the protein tetramer, 271 POPE (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine) lipids and 19769 (18443 for GlpF) TIP4P (168) water molecules. The GlpF structure was taken from the protein data bank (PDB code 1FX8 (13)). The starting structure of hAQP1 was modeled based on the x-ray structure of bovine AQP1 (PDB code 1J4N (15)) by mutating differing residues using the WHAT IF modeling software (167). Crystal water molecules were kept in the structures and chloride ions were added to neutralize the simulation systems. The OPLS all-atom force field (98, 99) was employed for the protein, lipid parameters were taken from Berger *et al.* (169). All simulations were equilibrated for 2 ns prior to production. A typical simulation box is shown in figure 4.2B.

The simulations were carried out using the GROMACS simulation software (106, 170). Electrostatic interactions were calculated every step with the particle-mesh Ewald method (114, 115). Short-range repulsive and attractive dispersion interactions were described together by a Lennard-Jones potential, which was cut off at 1.0 nm. The Settle (110) algorithm was used to constrain bond lengths and

angles of water molecules, and Lincs (109) was used to constrain all other bond lengths, allowing a time step of 2 fs. The simulation temperature was kept constant by weakly ( $\tau=0.1$  ps) coupling the protein, lipids, and solvent separately to a temperature bath (116) of 300 K. Likewise, the pressure was kept constant by weakly coupling the system to a pressure bath of 1 bar with a coupling constant  $\tau$  of 1 ps. During umbrella sampling simulations only, the box length in  $z$  direction (perpendicular to the membrane) was kept fixed.

The starting frames for the umbrella simulations were taken each from 20 ns equilibrium simulations of hAQP1 and GlpF. The aquaporin channels were divided into 0.25 Å wide equidistant sections parallel to the membrane with the center of each section representing an umbrella center. Subsequently, the solute was placed into the channel at the umbrella center. Water molecules that overlapped with the solute were removed. To enhance sampling, two to four solute molecules were placed in each pore at different positions. A distance between the solutes along the pore of at least 25 Å for water, ammonia, CO<sub>2</sub> and O<sub>2</sub>, and 30 Å for glycerol and urea was imposed to ensure that interactions between the solute molecules (if any) are negligibly small. During the subsequent energy minimization of the structure the protein RMSD was observed to make sure that the insertion of a larger solute into the pore did not lead to an unphysical widening of the pore.

Umbrella sampling calculations were carried out by applying a harmonic restraint force along the pore coordinate with force constants between 400 and 4000 kJ/mol/nm<sup>2</sup>. In ammonia the nitrogen atom was restrained, in water the oxygen, in CO<sub>2</sub>, urea and glycerol the central carbon, and in O<sub>2</sub> a dummy atom centered between the two oxygen atoms. Additionally, the solutes were restrained to a cylinder of radius  $r_c = 5$  Å whose axis was centered along the pore by applying an additional harmonic force  $F(r) = -k_c(r - r_c)H(r - r_c)$  pointing towards the cylinder axis. Here,  $r$  denotes the distance from the cylinder axis,  $k_c = 400$  kJ/mol/nm<sup>2</sup> the force constant, and  $H$  the Heaviside step function. The umbrella simulations of O<sub>2</sub>, CO<sub>2</sub>, NH<sub>3</sub> and water were carried out for 400 ps, the simulations of urea and glycerol for 600 ps and 1 ns, respectively.

Umbrella simulations for the lipid bilayers were performed by inserting the solute molecules at various positions into random snapshots taken from a 20 ns equilibrium run of a bilayer patch. The POPE and the POPC (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine) patch contained 128 lipid molecules each and 4777 or



5788 TIP4P water molecules, respectively. Up to 12 solute molecules were inserted into one simulation keeping a minimum distance of 25 Å to each other parallel and perpendicular to the bilayer. After energy minimization the simulations ran for at least 500 ps.

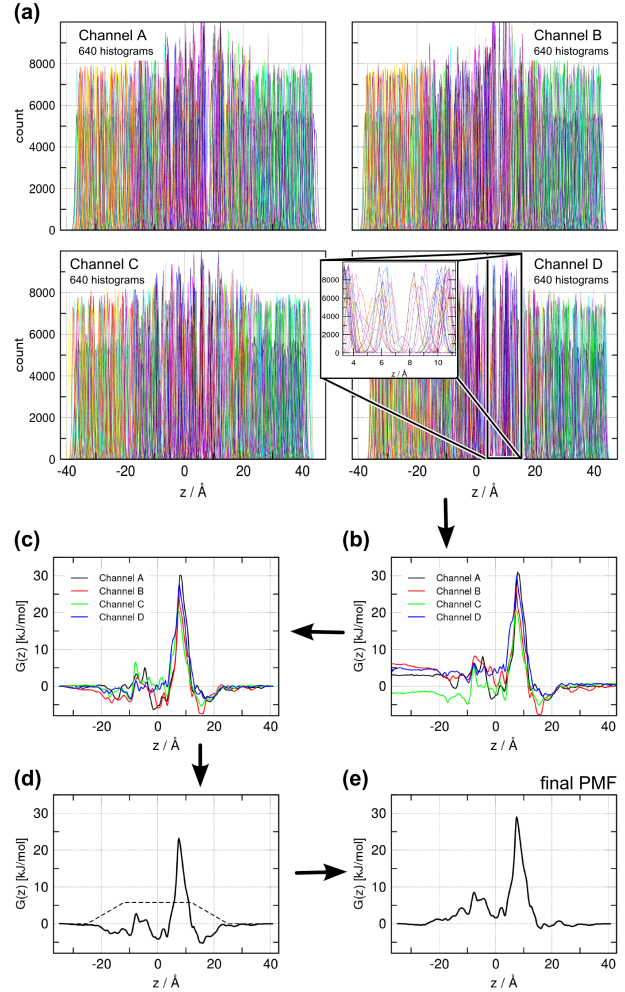
Ammonia parameters were taken from ref. (99), urea parameters from refs. (184, 185). Lennard-Jones parameters for O<sub>2</sub> were taken from the CHARM22 force field (104). The O<sub>2</sub> quadrupole was calculated from wavefunctions obtained at the MP2/6-311G\* level to  $Q_{zz} = -0.82 \text{ DÅ}$  and modeled by a positively charged dummy atom centered between two negatively charged oxygen atoms. CO<sub>2</sub> parameters were taken from ref. (181) and Glycerol was modeled from OPLS atom types for alcohols.

#### 4.2.2. Construction of PMFs

After removing the first 100 ps for equilibration, umbrella histograms 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 two NPA motifs of the corresponding channel. This procedure avoids a possible unphysical flattening of the PMF due to fluctuations of the monomers within the tetramer. It was ensured that all positions along the channel were well sampled. In case of poor sampling at maxima in the PMF additional umbrella simulations with higher force constants were performed. In total, 34432 histograms were collected from 1.6 μs of simulation of the aquaporin systems. The PMFs for permeation through the lipid bilayer were constructed from a total of 15232 histograms that were extracted from 695 ns of simulation.

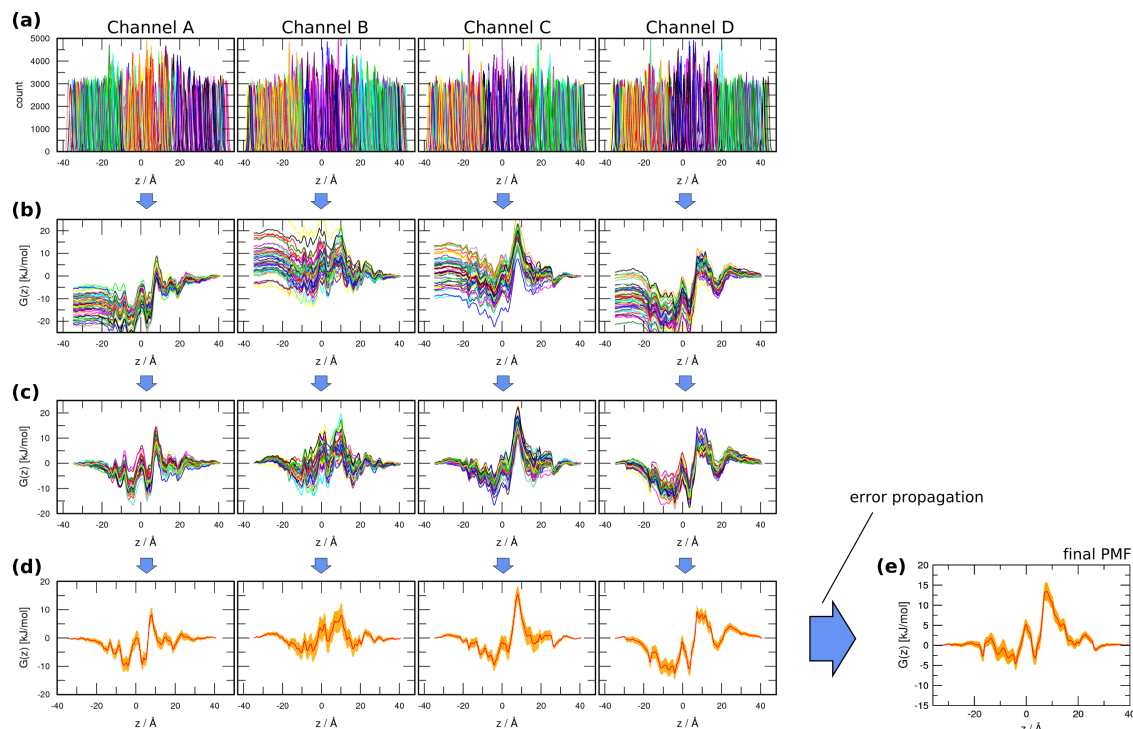
The construction of the PMFs illustrated in fig. 4.3. Combining the histograms of the four channels into an effective PMF turned out to be non-trivial. Side chains along the pore can adopt different conformations corresponding to a partial or even complete closure of the channel. In particular, the conserved Arg195 in hAQP1 can adopt several conformations, which has also been observed in the AQP-Z crystal structure (173). Such conformations are present in the starting structures of the umbrella runs. Opening and closure transitions are, however, not well sampled within an umbrella window of several hundred picoseconds. In such cases the weighted histogram analysis method (WHAM) (147) tends to overestimate the barrier when histograms from different channels are combined into a single WHAM

**Figure 4.3:** Example for the construction of a potential of mean force from umbrella simulations (here, for urea permeating through GlpF). **(a)** For each channel 640 histograms were extracted from umbrella simulations. **(b)** For each channel, the WHAM procedure was performed yielding four independent PMFs. **(c)** Possible offsets between the two bulk water regions were removed as described in the text. The procedure mimics the integration of the force from the two bulk regions to the point(s) of lowest sampling inside the channel. **(d)** The four channels were considered as independent pathways. Therefore, the four PMFs  $G_i(z)$  were combined to an effective PMF  $G_{\text{eff}}(z)$  via  $\exp(G_{\text{eff}}(z)/k_B T) = 4^{-1} \sum_{i=1}^4 \exp(G_i(z)/k_B T)$ . Due to the cylindrical constraint on the solute its entropy in bulk water is underestimated. In other words, the PMF relates to the area of the cylinder. To relate the PMF to the area of an aquaporin monomer a trapezoidal correction of 5.7 kJ/mol was employed in the entrance and exit region of the channel (dashed line), yielding the final PMF displayed in panel (e).



procedure. To alleviate this problem, the PMF was calculated for each monomer separately using the WHAM procedure. The four channels were considered as independent pathways. Therefore, the effective PMF  $G_{\text{eff}}(z)$  was calculated via  $\exp(-G_{\text{eff}}(z)/k_B T) = 4^{-1} \sum_{j=1}^4 \exp(-G_j(z)/k_B T)$ . Here,  $T$  denotes the temperature and  $k_B$  the Boltzmann constant.

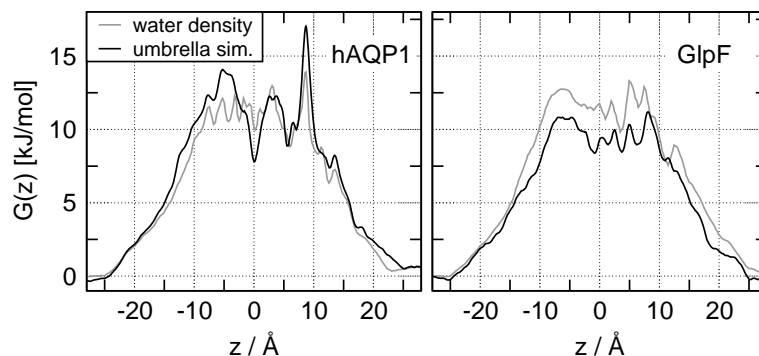
Prior to combining the four profiles into  $G_{\text{eff}}(z)$ , the single channel PMFs were corrected for a possible free energy offset between the two bulk water regions. The offset was distributed between adjacent bins proportional to  $1/(n_i n_{i+1})^\alpha$ , where  $n_i$  and  $n_{i+1}$  are the number data points in the adjacent bins  $i$  and  $i+1$  as collected from all histograms of the channel. A reasonable value was found to be  $\alpha = 2$  as it yields a PMF referring to an integration of the mean force from the two bulk water regions to the point(s) of lower sampling inside the pore, typically the barrier at the conserved arginine. The exact choice of  $\alpha$  was found to have only minor



**Figure 4.4.:** Example for the determination of the statistical uncertainty for potentials of mean force (PMF) using bootstrap analysis (here, for  $\text{CO}_2$  permeating through GlpF). In order to account properly for limited long time sampling bootstrap analysis was performed by considering only complete histograms as independent data points, yielding a rather conservative estimation for the uncertainty. First, bootstrap analysis was performed for each channel separately as displayed in the four columns. From  $N$  histograms available [panel (a)],  $N$  histograms were randomly selected (allowing duplication) and used in a subsequent weighted histogram analysis method (WHAM). Hence, we considered only complete histograms as independent data points yielding a rather conservative estimation for the uncertainty. It was made sure that the  $N$  selected histogram covered the complete  $z$  range and that no gaps were generated. Bootstrapping was done 50 times yielding 50 PMFs for each channel [panel (b)]. The constraint that the PMF equals in the two bulk water regions was incorporated as described in the methods [panel (c)]. From the 50 PMFs, the average and standard deviations were calculated denoted by the red lines and orange area in panel (d). The final PMF was calculated via  $\exp(-G_{\text{eff}}(z)/k_B T) = 4^{-1} \sum_{i=1}^4 \exp(-G_i(z)/k_B T)$  (see methods) allowing to employ error propagation from textbooks to calculate the uncertainty of the final PMF. It is displayed as an orange area around the final PMF (red line) in fig. (e). The uncertainty was generally  $\leq 2.5$  kJ/mol.

influence on the main barrier height ( $< 2$  kJ/mol).

Since the solutes were constricted to a cylinder along the pore the umbrella simulations yield the free energy profile referring to the cross section of the cylinder ( $A_c$ ). However, in order to render the profile comparable to the profile for permeation through the lipid bilayer (which is independent of the area), it should



**Figure 4.5.:** Potentials of mean force for water permeating through human aquaporin-1 (hAQP1, left) and GlpF (right) as determined from the water density in equilibrium simulations (shaded lines) and by umbrella sampling calculations (black lines). The two methods agree within 3 kJ/mol.

refer to the membrane area occupied by the aquaporin monomer ( $A_{\text{mono}}$ ) for which the profile has to be corrected. The correction of this purely entropic effect reads  $\Delta G = k_B T \ln(A_{\text{mono}}/A_c)$ . We estimated  $A_{\text{mono}}$  to 10.3 nm<sup>2</sup> and 10.7 nm<sup>2</sup> for hAQP1 and GlpF, respectively. From the cylinder radius  $r_c$  and the force constant  $k_c$ ,  $A_c$  can be estimated to  $A_c = 1.05$  nm<sup>2</sup> yielding a correction of 5.7 kJ/mol and 5.8 kJ/mol for hAQP1 and GlpF, respectively. This correction was incorporated in the PMFs as a linear trapezoidal correction in the pore entrance and exit region.

**Uncertainty estimation using bootstrap analysis** Statistical errors were calculated using bootstrap analysis. The procedure is illustrated in fig. 4.4 and described in the figure caption. The uncertainty at the main barriers was  $\leq 2.5$  kJ/mol in the aquaporin channels and  $\leq 2$  kJ/mol in the lipid bilayer systems.

**Validation** To validate the umbrella sampling calculations the PMFs for water permeating through hAQP1 and GlpF were independently calculated from the water density in the equilibrium trajectories. Figure. 4.5 shows the PMFs for hAQP1 and GlpF determined by umbrella sampling and calculated from the water density. The two methods agree within 3 kJ/mol.

**Direct interactions** Water-protein and solute-protein interactions (fig. 4.10) were calculated as the sum of Lennard-Jones and short-range coulomb interactions. Only configurations that contribute to permeation events were taken into account. There-

fore, snapshots were evaluated only when the distance between Arg195 (Arg206 in GlpF) and the opposite His180 (Phe200 in GlpF) was not reduced by more than 1 Å compared to the crystal structures (13, 15). The analysis of arginine interactions was restricted to arginine atoms along the pore.

## 4.3. Results

### 4.3.1. Solute permeabilities

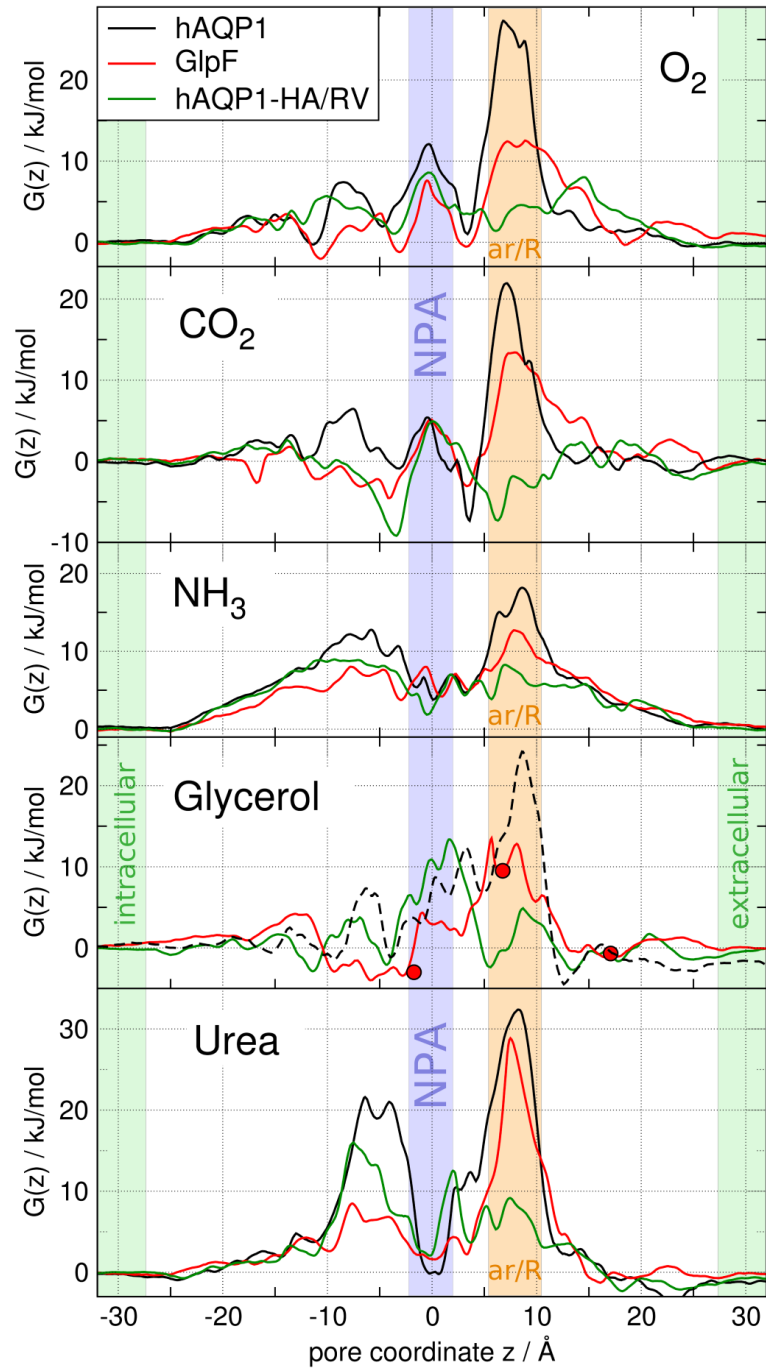
Figure 4.6 presents the PMFs for permeation of O<sub>2</sub>, CO<sub>2</sub>, NH<sub>3</sub>, glycerol and urea permeating through hAQP1 (black curves) and GlpF (red curves). The PMFs  $G(z)$  are shown as a function of the pore coordinate  $z$  where  $z = 0$  corresponds to center of mass of the conserved NPA motifs.

Figure 4.6 shows that the main barrier for each solute permeating through hAQP1 or GlpF is located in the ar/R region demonstrating its role as the selectivity filter for uncharged solutes. For all solutes, the barrier in GlpF at the ar/R site is lower than the barrier in hAQP1. Hence, GlpF is more permeable to the solutes and is less selective than hAQP1. In contrast to the ar/R region, the highly conserved NPA motifs are not involved in the selectivity for uncharged solutes as they form only a minor barrier against permeation of the rather apolar O<sub>2</sub> and CO<sub>2</sub> molecules. The height of the main barriers are summarized in table 4.1.

To address the physiological relevance of permeation through AQPs the barriers were compared to barriers against the permeation across the lipid bilayer. Figure 4.7 shows the PMFs for permeation of urea, glycerol, H<sub>2</sub>O, NH<sub>3</sub>, CO<sub>2</sub> and O<sub>2</sub> through two phospholipid membranes, one composed of pure POPE and the other of pure POPC. The highest barrier for the apolar O<sub>2</sub> and CO<sub>2</sub> molecules permeating through either POPE or POPC is  $\leq 6$  kJ/mol indicating that these membranes are highly permeable to small apolar molecules such as O<sub>2</sub> and CO<sub>2</sub> (fig. 4.7, green and yellow curves). The barriers against O<sub>2</sub> and CO<sub>2</sub> permeation through hAQP1 and GlpF are significantly higher than the corresponding membrane barriers. Hence, these aquaporin channels embedded in membranes similar to POPE or POPC are not expected to increase O<sub>2</sub> or CO<sub>2</sub> flux across the membrane.

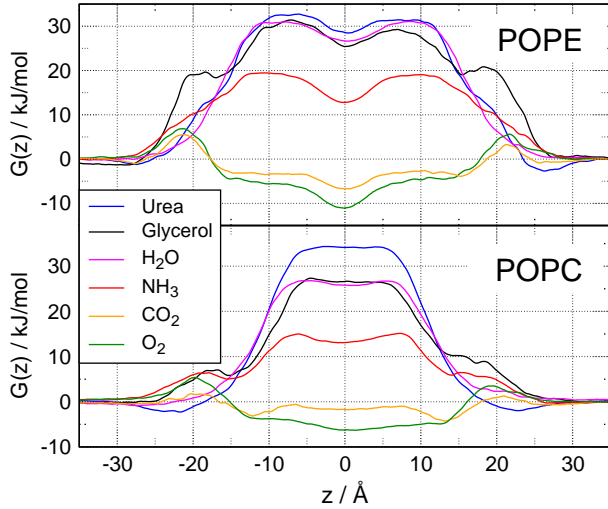
The barriers against ammonia permeation through POPE (19 kJ/mol) or POPC (15 kJ/mol) are considerably higher than the barriers for apolar gas molecules reflecting its more hydrophilic nature. Compared to O<sub>2</sub> or CO<sub>2</sub>, these barriers refer to





**Figure 4.6.:** Potentials of mean force  $G(z)$  for  $O_2$ ,  $CO_2$ ,  $NH_3$ , glycerol and urea (from top to bottom) permeating through the monomeric water pores of human aquaporin-1 (black curves), GlpF (red) and the H180A/R195V mutant of aquaporin-1 (green).  $z=0$  corresponds to the center of the NPA motifs. The NPA region is highlighted by a blue bar, the aromatic/arginine constriction region by an orange bar. The red circles correspond to the glycerol positions in the GlpF crystal structure (13).





**Figure 4.7:** Potentials of mean force for permeation of  $O_2$ ,  $CO_2$ ,  $NH_3$ ,  $H_2O$ , glycerol, and urea through a membrane of pure POPE (top) and POPC (bottom).  $z=0$  corresponds to the membrane center.

a permeability reduced by a factor of 80-400. Here, we assumed the permeability to be proportional to  $\exp(-\Delta G_{\max}/k_B T)$  where  $\Delta G_{\max}$  denotes the maximum barrier height. When permeating through hAQP1 and GlpF, ammonia encounters barriers of 18 kJ/mol and 12.5 kJ/mol, respectively, indicating that hAQP1 is unlikely to increase  $NH_3$  permeation whereas GlpF can significantly enhance the  $NH_3$  flux across the membrane. This observation is in line with an experimental study on ammonia permeation across the oocyte membrane (60). In that study, no influence of AQP1 on  $NH_3$  flux was measured whereas aquaglyceroporins such as AQP3 and AQP9 were found to increase  $NH_3$  permeation.

Both, glycerol and urea, encounter substantial barriers between 27 and 34 kJ/mol against permeation through POPE or POPC underlining the need for membrane

**Table 4.1.:** Summary of free energy barriers  $\Delta G_{\max}$  in kJ/mol

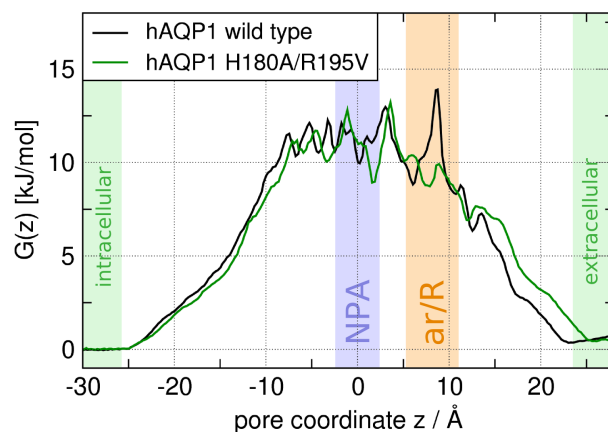
Solute	hAQP1	GlpF	hAQP1- HA/RV	POPE	POPC
$O_2$	27	12.5	8.5	6	4
$CO_2$	22	13.5	5	4	1.5
$NH_3$	18	12.5	9	19	15
$H_2O$	14*	13.5*	13*	31	27
Glycerol	(24)	13.5	13.5	30	27
Urea	32.5	29	16	32	34

\*Determined from water density. All other barriers from umbrella sampling simulations.

channels if a substantial glycerol or urea flux is required by the metabolism. As expected, GlpF is permeated efficiently by glycerol with a maximum barrier of only 13.5 kJ/mol. This value is significantly lower than the activation energy of 30.5 kJ/mol reported in a previous MD study (134). Compared to membranes of pure POPE or POPC, the barrier of 13.5 kJ/mol results in an increase in the permeability by a factor of 750 or 230, respectively, when assuming dense expression of GlpF. These values compare favorably with the reported increase by a factor of 400 for proteoliposomes reconstituted with GlpF (64). In contrast, urea encounters a large barrier of 29 kJ/mol against permeation through GlpF confirming that GlpF is not an efficient urea channel. Solvating GlpF in a membrane of POPE or POPC is expected to increase the permeability for urea only 3-fold or 7-fold, respectively, in agreement with only 3-fold increase measured for liposomes (64). As expected, hAQP1 does not enhance urea flux across the membrane as its ar/R region forms a barrier against urea permeation well above 30 kJ/mol.

After inserting glycerol into the ar/R region of hAQP1 we observed a widening of the pore during the energy minimization (EM) of the structure. The widening was visible from the RMSD drift of the protein atoms during the EM and did not occur in any other simulation (data not shown). These findings confirm that the ar/R site of hAQP1 is too narrow to allow passage of glycerol. The unphysical widening of the pore lead to an underestimation of the barrier of 24 kJ/mol for glycerol permeation through hAQP1 (fig. 4.6, dashed line). The true barrier is likely to be substantially higher, in line with experiments that did not observe glycerol permeation through AQP1 (186).

The role of the ar/R region as the selectivity filter for uncharged solutes can be further investigated by point mutations in the ar/R region. Here, we studied the effect of the H180A/R195V double mutant of hAQP1 (hAQP1-HA/RV) since H180 and R195 are generally considered as key residues for the selectivity of AQP1. A recent experimental study investigated permeation properties of several hAQP1 mutants including hAQP1-HA/RV (183). The authors found some of the mutations to enhance  $\text{NH}_3$  permeability or even to allow urea or glycerol permeation through hAQP1. To further rationalize these findings, fig. 4.6 shows the PMFs for  $\text{O}_2$ ,  $\text{CO}_2$ ,  $\text{NH}_3$ , glycerol and urea permeating through hAQP1-HA/RV as green curves. For all solutes, the barrier in the ar/R region is drastically reduced by the mutation emphasizing the importance of the ar/R residues for channel selectivity. In con-



**Figure 4.8.:** PMFs for water permeating through human aquaporin-1 (hAQP1, black curve) and the hAQP1-H180A/R195V mutant (green curve). Although the mutation at the ar/R site renders the region more hydrophobic the mutant shows an even lower maximum barrier than the hAQP1 wild type. Hence, the mutant is highly permeable to water, in agreement experimental findings (183). The PMFs have been computed from the water density in equilibrium simulations.

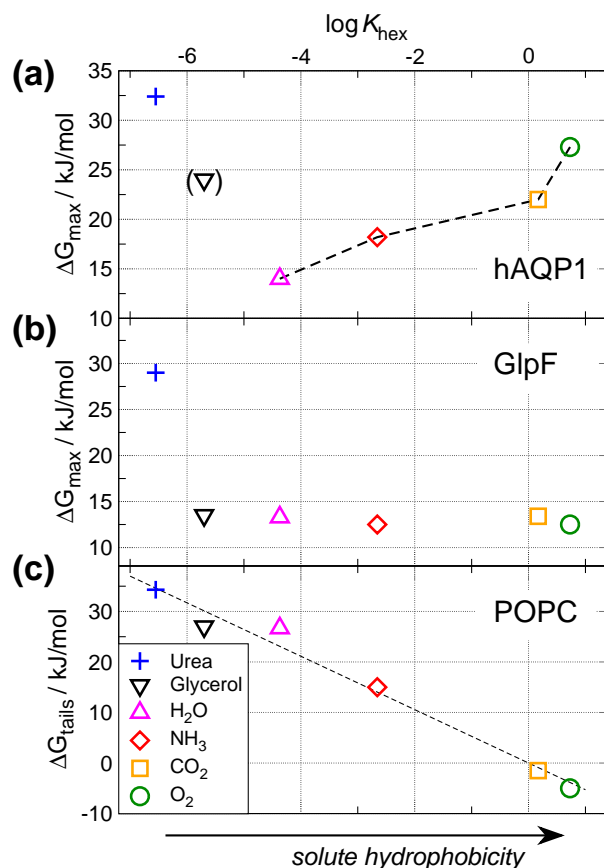
trast to the wild type, the main barriers are not located in the ar/R region any more.  $O_2$  and  $CO_2$  can almost freely diffuse through hAQP1-HA/RV. In agreement with ref. (183), the mutation renders hAQP1 an efficient  $NH_3$  channel and allows passage of urea whereas water permeability is hardly affected (compare fig. 4.8 and table 4.1). In addition, our results indicate that glycerol is likely to permeate through hAQP1-HA/RV, although ref. (183) reported glycerol permeation through the hAQP1-F56A/H180A mutant only.

### 4.3.2. Solute hydrophobicity

Figure 4.9a displays the height of the main barrier for urea, glycerol,  $H_2O$ ,  $NH_3$ ,  $CO_2$ , and  $O_2$  against permeation through hAQP1 as a function of solute hydrophobicity. The solute hydrophobicity is measured as the logarithm of the hexadecane/water partition coefficient  $\log K_{\text{hex}}$  of the solute at 298 K.  $K_{\text{hex}}$  values were taken from references (149, 150, 151, 152). For the small solutes  $H_2O$ ,  $NH_3$ ,  $CO_2$ , and  $O_2$  the maximum barrier clearly increases with the solute hydrophobicity. The larger and very hydrophilic glycerol and urea molecules are not excluded by hydrophobicity but rather sterically. Hence, the ar/R region of hAQP1 can be considered as both a hydrophobicity filter and a size filter.

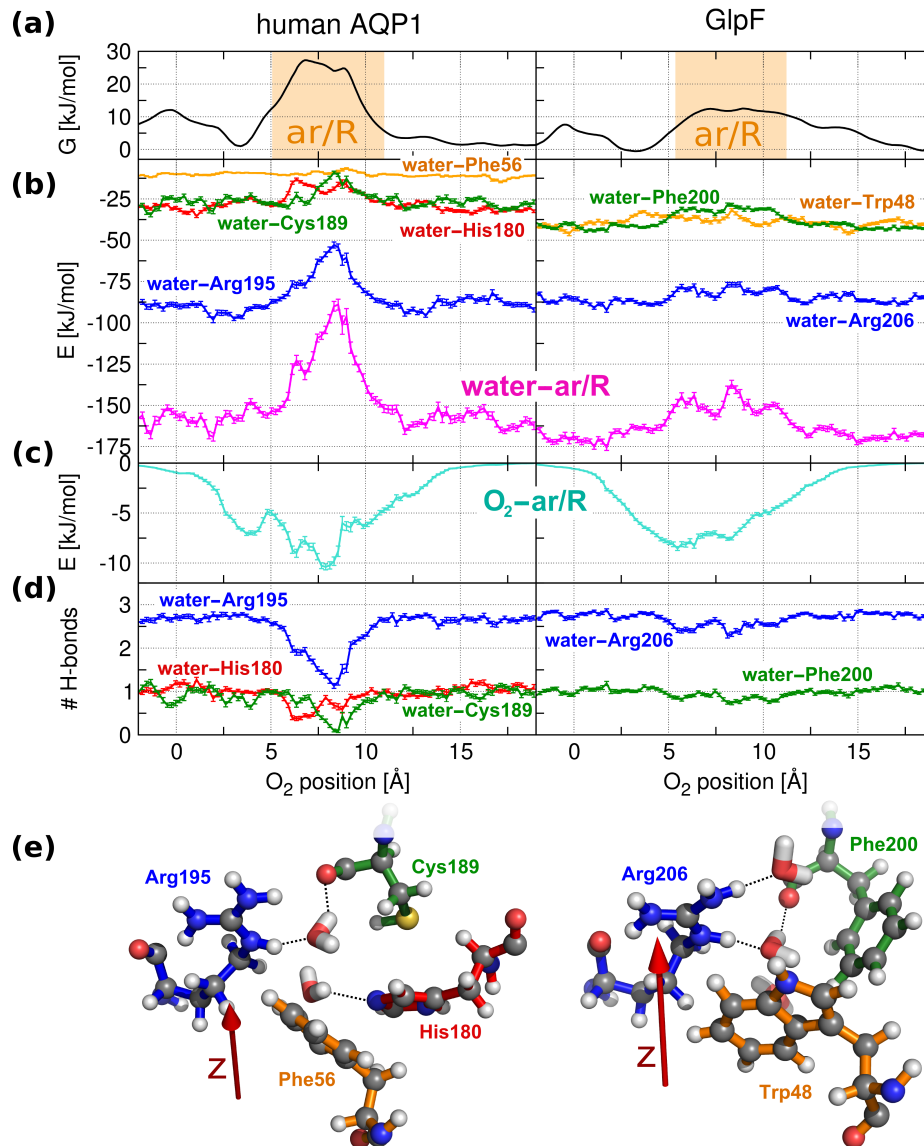
Figure 4.9b shows the corresponding plot for GlpF. The larger and more hy-

**Figure 4.9:** Permeability as a function of hydrophobicity of the solute. (a) and (b) Free energy barrier  $\Delta G$  for urea (plus), glycerol (triangle down),  $H_2O$  (triangle up),  $NH_3$  (diamond),  $CO_2$  (square) and  $O_2$  (circle) permeating along the pores of aquaporin-1 (a) and GlpF (b).  $\Delta G$  is plotted versus the logarithm of the hexadecane-water partition coefficient  $\log K_{hex}$  of the solute which is a common measure for hydrophobicity. (c) Solvation free energy difference between the solute in water and the solute in the hydrophobic tail region of POPC as determined by umbrella simulations. The dashed line is a linear fit to the data points to guide the eye.



drophobic ar/R region of GlpF is less selective than the ar/R region of hAQP1 forming a significant barrier only against permeation of urea but efficiently allowing the passage of glycerol. The difference between figs. 4.9A and 4.9B demonstrates that the differing residues in the ar/R regions of hAQP1 and GlpF account for highly different permeation characteristics.

For comparison, fig. 4.9c shows the energetic cost  $\Delta G_{tails}$  for moving the solute from bulk water to the hydrophobic lipid tail region of a POPC membrane. As expected, the bilayer forms a filter permeated by hydrophobic molecules. The linear dependence between  $\Delta G_{tails}$  and  $\log K_{hex}$  favorably compares to Overton's rule stating that the permeability of a membrane with respect to different solutes is approximately proportional their oil/water partition coefficients. Hence, fig. 4.9C validates the simulation parameters and protocol.



**Figure 4.10.:** Water-protein interaction as selectivity mechanism for aquaporins. Comparison between human aquaporin-1 (hAQP1, left panel) and GlpF (right panel). **(a)** Potentials of mean force for  $O_2$  permeating through the aromatic/arginine regions of hAQP1 and GlpF versus the position of  $O_2$ . **(b)** Interactions between water and the aromatic/arginine (ar/R) residues as a function of the position of a permeating  $O_2$  molecule (magenta curves). In AQP1, water-ar/R interactions are reduced by  $\sim 60$  kJ/mol when a  $O_2$  molecule is present in the ar/R region, and the loss of water-protein interaction cannot be compensated by  $O_2$ -protein interactions which are displayed in panel (c). In GlpF, water-ar/R interactions are hardly affected by the  $O_2$ . The water-ar/R interaction can be decomposed into interactions between water and single residues (blue, orange, green and red curves, as denoted in the figure). **(d)** Number of hydrogen bonds between water and protein residues versus the position of a permeating  $O_2$  molecule. **(e)** MD snapshots of the ar/R regions of hAQP1 and GlpF including several water molecules. The residues are colored according to the curves in (b) and (d). Possible water-protein hydrogen bonds are denoted by dashed lines. The red arrow indicates the pore coordinate.

### 4.3.3. Water–channel interactions as selectivity mechanism

As demonstrated in the previous section, the ar/R region of hAQP1 is a filter permeated by small polar molecules. A previous MD study showed that the barrier that CO<sub>2</sub> encounters in the ar/R region is not due to unfavorable CO<sub>2</sub>–protein interactions (181). Instead, water–Arg195 hydrogen bonds were observed to open upon CO<sub>2</sub> passage. Here, we investigate the selectivity mechanism of hAQP1 in detail. As an example, we consider a permeation event of the apolar O<sub>2</sub> molecule (fig. 4.10). To guide the eye, fig. 4.10A shows the PMFs for O<sub>2</sub> permeating through hAQP1 and GlpF, respectively. Figure 4.10B presents the interaction between water and key residues of the ar/R region versus the position of a O<sub>2</sub> molecule inside the channel. In the left panel, water–protein interactions in hAQP1 are presented, in the right panel the interactions in GlpF. Likewise, fig. 4.10C displays the interaction between the O<sub>2</sub> molecule and the ar/R residues as a function of O<sub>2</sub> position and fig. 4.10D the average number of water–protein hydrogen bonds. In fig. 4.10E, MD snapshots of the ar/R regions of hAQP1 and GlpF are shown in ball and stick representation. The residues are colored according to the curves in the graph, and possible water–protein H-bonds are indicated by dashed lines.

The key finding is that upon O<sub>2</sub> passage through hAQP1 favorable interactions between water and the ar/R residues (magenta curve) are reduced by  $\sim 60$  kJ/mol. The loss of water–protein interaction cannot be compensated by O<sub>2</sub>–protein interaction since the apolar O<sub>2</sub> interacts with the ar/R residues in the order of 10 kJ/mol only. The reduction in water–protein interaction occurs right at the barrier in the PMF. Decomposing the ar/R in the single residues (blue, orange, green, and red curve for Arg195, Phe56, Cys189, and His180, respectively) shows that the water–protein interaction is mainly affected by reduced interaction to Arg195 ( $\sim 35$  kJ/mol). In addition, interaction to Cys189 and His180 are reduced by  $\sim 15$  kJ/mol each. In the wider and more hydrophobic ar/R region of GlpF, water–ar/R interactions are hardly affected by the position of a passing O<sub>2</sub> molecule (fig. 4.10B, right panel). Hence, small solutes pass the ar/R region of GlpF without reducing water–protein interactions. The same trend can be observed in fig. 4.10D, which presents the average number of water–protein hydrogen bonds as a function of O<sub>2</sub> position. Upon O<sub>2</sub> passage through hAQP1, more than one H-bond between water and Arg195 is lost (blue line). In addition, the probability for an H-bond between water and both His180 or Cys189 is significantly reduced (red



and green lines). In contrast, water–Arg206 H-bonds in GlpF are hardly affected by a permeating O<sub>2</sub> molecule.

## 4.4. Discussion

From the extensive set of MD simulations presented here a detailed understanding of the selectivity mechanism of aquaporins can be derived. We showed that in both aquaporins and aquaglyceroporins the ar/R region is the selectivity site for uncharged solutes. In hAQP1, the narrow and hydrophilic ar/R site forms a filter permeated by small polar solutes. Upon passage of a solute, favorable interactions between water and protein (mainly to Arg195, but also to Cys189 and His180) are reduced and need to be replaced by solute–protein interactions. The more polar the solute, the stronger it can interact with hydrophilic ar/R residues rendering a lower energetic cost to replace a water molecule. Larger solutes such as urea or glycerol are sterically excluded in hAQP1.

The hAQP1-HA/RV double mutant was found to be highly permeable to water with an even lower barrier against water permeation compared to the hAQP1 wild type (fig. 4.8). Hence, Arg195 and His180 are not necessary to isolate single water molecules from the bulk, which is in agreement with experiments that measured high water flux across the hAQP1-HA/RV mutant (183). We showed that these residues are, however, important for the selectivity for uncharged solutes since mutations in the ar/R region have drastic effects on the channel characteristics. The mutations result in a larger and more hydrophobic pore, thus eliminating the selectivity mechanism for small polar solutes, which is present in the hAQP1 wild type. Without the selectivity mechanism and due to the larger pore size the hAQP1-HA/RV mutant is highly permeable to apolar solutes as well as to NH<sub>3</sub> and urea.

In aquaglyceroporins the ar/R region is wider than in AQP1 and has a more hydrophobic pocket opposite to the arginine. In GlpF this environment allows passage of polyols whose hydroxyl groups can well interact with Arg206 replacing water–Arg206 H-bonds whereas its apolar backbone matches into the hydrophobic pocket (13). Our results demonstrate that due to the wider and more hydrophobic ar/R site, the selectivity mechanism present in AQP1 does not apply in aquaglyceroporins similar to GlpF. Therefore, GlpF-like aquaglyceroporins are efficiently

permeated by small solutes such as  $\text{NH}_3$ ,  $\text{CO}_2$  or  $\text{O}_2$ .

We stress that from the permeabilities of the wide range of solutes that were probed in the present study the permeability of AQP1 and GlpF for other small solutes may be estimated. The results suggest that small hydrophobic solutes such as  $\text{NO}$ ,  $\text{CO}$ , or  $\text{N}_2\text{O}$  are unlikely to permeate through AQP1 with barriers well above 20 kJ/mol. In contrast, aquaglyceroporins similar to GlpF are expected to be well permeable to such solutes.

To address the possible flux of molecules such as  $\text{CO}_2$  or  $\text{NO}$  through AQP1 we suggest permeation experiments on GlpF or the hAQP1-HA/RV mutant. If permeation occurs along the AQP1 water channels and not through the lipid bilayer the different ar/R region should drastically increase the permeability. For GlpF, for example, we would expect an 30-fold increase of the  $\text{CO}_2$  flux compared to hAQP1.

## 4.5. Conclusions

A simple filter mechanism governs the permeability of AQP1. For small solutes permeating through AQP1, a hydrophobic effect leads to a remarkable anti-correlation between permeability and hydrophobicity. Large solutes are sterically excluded. Hence, the ar/R region of AQP1 is a filter against both hydrophobicity and size. This filter mechanism does not apply in GlpF that efficiently permeates all investigated solutes except for urea. Surprisingly, not solute-pore but water-pore interactions complemented by steric effects emerged as the key determinants underlying the selectivity mechanism and the barrier height in aquaporins and aquaglyceroporins.



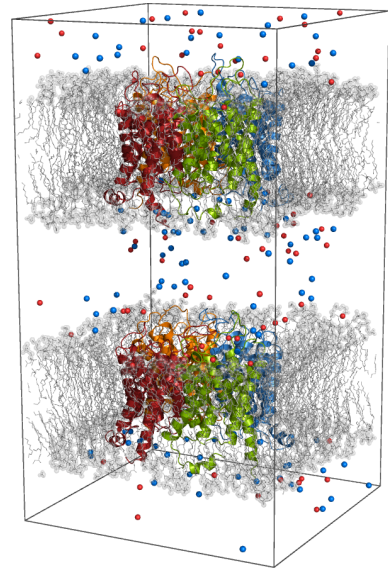
## 5. Is aquaporin-1 a voltage-regulated water channel?

Regulation of protein function is important for any living organism. The cell can adjust the protein function by up- or down-regulating the protein expression, or by a direct activation or inhibition of proteins which have already been expressed. Such a direct regulation may be triggered by

- pH, affecting the protonation state of some residues,
- binding or unbinding of a ligand or an ion,
- phosphorylation of one or several protein residues,
- glycosylation of an asparagine, serine, or threonine residue,
- temperature,
- the electrostatic membrane potential which triggers the gating of ion channels, a mechanism which is crucial for the function of neurons.

Plant aquaporins are known to be regulated by a number of conditions which possibly include phosphorylation, heteromerization, pH,  $\text{Ca}^{2+}$  ions, pressure, solute gradients and temperature (187). For example, the spinach plasma membrane aquaporin SoPIP2;1 was proposed to be gated by (de)phosphorylation of two serine residues at the cytoplasmic side the the channel (16, 63, 87, 188).

Here, we investigate the influence of an electrostatic membrane potential on the water flux of aquaporin-1 (AQP1). We employ extensive MD simulations of AQP1 subject to membrane potentials between  $-550$  and  $+550$  mV. The simulations reveal a significant impact of the membrane voltage on channel permeability and suggest a possible regulatory mechanism.



**Figure 5.1:** Simulation box of two aquaporin-1 tetramers (cartoon representation) embedded in POPE membranes (grey) and solvated in water with physiological salt concentration. Cations are shown as red and anions as blue spheres. For clarity, the water is not shown.

## 5.1. Methods

**Simulation setup with a membrane potential** A equilibrated simulation box of hAQP1 was chosen as starting configuration for all simulations. The box was set up as explained in section 3.1.2. Two copies of the simulation box were stacked on top of each other, creating a system of two AQP1 tetramers, each solvated in a membrane of POPE (fig. 5.1). Hence, two separated compartments of water were generated by the procedure. A physiological salt concentration of sodium chloride (150mM) was added to the water compartments.

An electrostatic potential across the two membranes  $\Delta V$  was generated by adding few additional cations into the central water compartment and additional anions into the water compartment at the top and the bottom of the simulation box. To allow the increase of  $\Delta V$  in small steps artificial ions with fractional charges of  $\pm e/2$  or  $\pm e/4$  were added if necessary. Here,  $e$  denotes the unit charge. In total, 13 simulation systems were generated with a maximum additional charge in the water compartments of  $\pm 3e$ , in steps of  $0.25e$ .

After energy minimization each simulation system ran for 20 ns, omitting the first 2 ns for equilibration. All forcefield and simulation parameters were chosen as explained in section 3.1.2.

**Calculation of  $p_f$**  The single-channel water permeability  $p_f$  is defined via the net molar water flux  $\Phi$  across a channel in response to an osmotic gradient,  $\Phi =$



$p_f \Delta C_s$ . Here,  $\Delta C_s$  denotes the difference in concentration of a non-permeating solute between the two water reservoirs connected by the channel (189).  $p_f$  was derived from the simulations as follows. The narrowest part of the pore including the NPA region and the aromatic/arginine (ar/R) constriction region (compare fig. 3.3) was split into three sections of thickness  $3.8 \text{ \AA}$ , referring to a typical water-water distance inside the pore. During the simulation, the number of permeation events across each section was counted and subsequently averaged over the three sections. In case of the presence of a single-file water column the permeation events across the sections refer to a displacement of the water column inside the channel by one water-water distance, or, in case of a broken water file, to water motions which partially contribute to a permeation event. The number of such water hops per time ( $\omega_w$ ) relates to  $p_f$  via  $p_f = V_w \omega_w / 2$  (190, 191). Here,  $V_w = 18 \text{ cm}^3/\text{mol}$  denotes the molar volume of water. Statistical errors were estimated by binning analysis. For this purpose, the trajectories were split into pieces of 3 ns each which were subsequently analyzed with respect to  $p_f$ . The  $p_f$  values as collected from the bins were subsequently used to compute the confidence intervals (compare fig. 5.3).

**Membrane potential** The electrostatic potential  $\Phi(\mathbf{r})$  is related to the charge density  $\rho(\mathbf{r})$  via the Poisson equation  $\Delta \Phi = 4\pi\epsilon_0\epsilon_r\rho$ . Here,  $\Delta$  denotes the Laplace operator. In the simulation, the permittivity  $\epsilon_r$  equals unity since the polarization of the medium is already explicitly included in  $\rho(\mathbf{r})$ . Integrating out the  $x$ - $y$  degrees of freedom, i.e. averaging  $\Phi$  and  $\rho$  in the  $x$ - $y$  plane leads to the one-dimensional Poisson equation

$$\frac{\partial^2 \Phi(z)}{\partial z^2} = 4\pi\epsilon_0\rho(z). \quad (5.1)$$

Hence,  $\Phi(z)$  can be computed by twice integrating  $\rho(z)$ ,

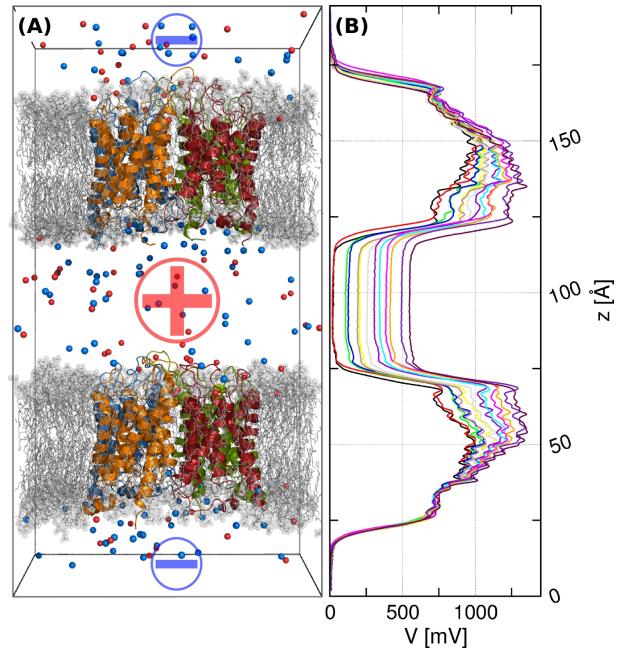
$$\Phi(z) = 4\pi\epsilon_0 \int_0^z dz' \int_0^{z'} dz'' \rho(z''). \quad (5.2)$$

## 5.2. Results and discussion

### 5.2.1. Single channel water permeability $p_f$

Figure 5.2B displays the electrostatic potential  $\Phi(z)$  across the box as calculated from the 13 different 20-ns simulations. The additional ions generated membrane

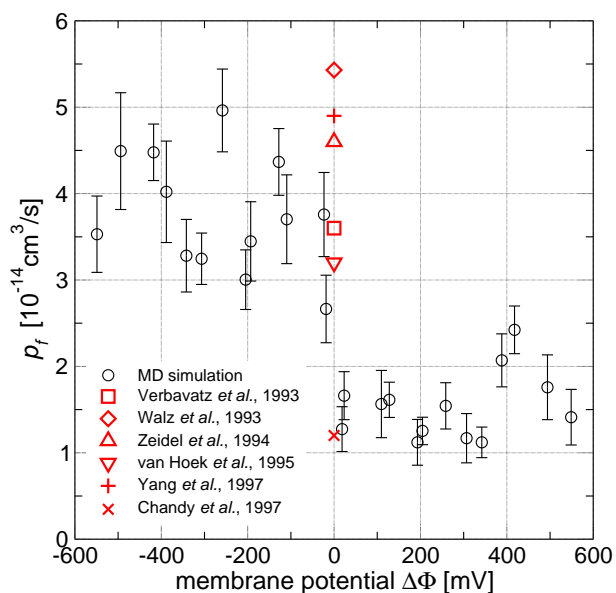
**Figure 5.2:** (A) Simulation box of two aquaporin-1 tetramers embedded in membranes and solvated in water at physiological salt concentration (150mM). To generate an electrostatic potential across the membranes some cations were additionally added to the central water reservoir, and some anions were added to the water reservoir at the bottom and the top of the figure (as indicated by the ‘+’ and ‘-’ symbols). (B) Electrostatic potential  $\Phi(z)$  versus the box coordinate  $z$  (perpendicular to the membranes) as derived from the atom positions during the 13 simulations. Note the potential difference between the central water compartment and the compartment at the top and bottom the box.



potentials  $\Delta\Phi$  between 0 and +550 mV. Hence, the upper AQP tetramer was subjected to membrane potentials between  $-550$  and 0 mV and the lower AQP tetramer to membrane potentials between 0 and +550 mV.

The single-channel permeability  $p_f$  as derived from the simulations is plotted in figure 5.3 as a function of the membrane potential  $\Delta\Phi$ .  $p_f$  shows a clear dependence on  $\Delta\Phi$  with a step around  $\Phi \sim 0$  mV.  $p_f$  values at negative potential (with the electric field pointing in positive  $z$ -direction; upper tetramer in fig. 5.2) are approximately two times larger than  $p_f$  values at positive potential (with the electric field pointing in negative  $z$ -direction; lower tetramer in fig. 5.2).

Experimental studies during the last years reported different values for  $p_f$  of AQP1, ranging from  $1.2$  to  $11.7 \times 10^{-14}$  cm<sup>3</sup>/s (192). Such experiments are challenging as the calculation of  $p_f$  requires knowledge on the density of channels in the membrane. Figure 5.3 shows the experimental  $p_f$  taken from refs. (28, 193, 194, 195, 196, 197). Most experiments agree at least qualitatively with the simulated  $p_f$  at negative membrane potential. This could be taken as an indication the shift in  $p_f$  due the potential may not be located  $\Delta\Phi \approx 0$  mV as predicted from the applied force field, but rather at positive  $\Delta\Phi$ .



**Figure 5.3:** Single channel permeability  $p_f$  as a function of the membrane potential (circles). The experimental  $p_f$  at zero potential as denoted in the figure legend were taken from references (28, 193, 194, 195, 196, 197).

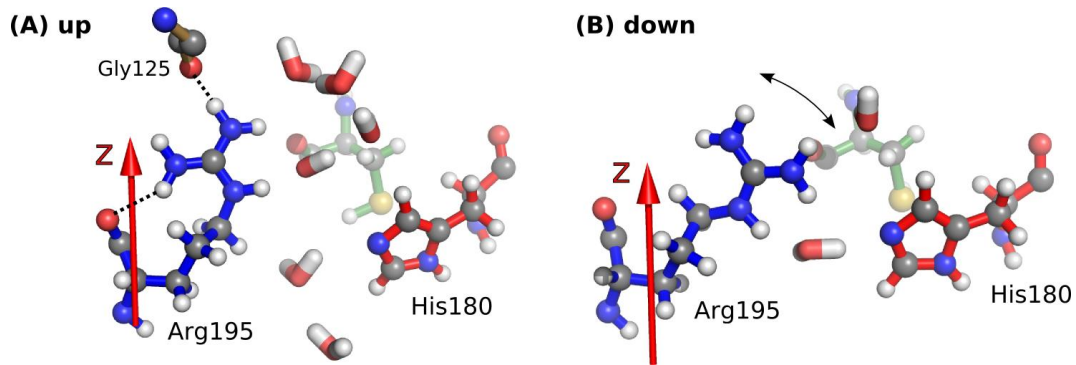
## 5.2.2. A putative regulatory mechanism

### The two states of Arg195

During all simulations of AQP1 the conserved Arg195 turned out to highly flexible adopting several distinct configurations within its surroundings. In particular, two states of the dihedral angle around the  $C_\gamma-C_\delta$  bond were frequently visited. The two dihedral states, an *up* and a *down* state of Arg195, are visualized in figure 5.4. Figure 5.4A shows the dihedral state of Arg195 that has been found in the crystal structure of both bovine and human AQP1 (15, 198). Figure 5.4B displays the second dihedral state which is frequently present during the simulations. Transitions between the *up* and the *down* state occurred typically on time scales of few nanoseconds. It is interesting to consider whether this effect is dependent on the applied force field. For the present work we mainly applied the OPLS all-atom force field. However, the two distinct dihedral states were also visited in related AQP1 simulations that used the GROMOS96 force field and have been reported in simulations applying the CHARMM27 force field (199, 200). Additionally, in the *E. coli* water channel AQP-Z the two dihedral states of the conserved arginine have been observed in both the crystal structure (173) and in MD simulations (182).

In the *up* state of Arg195 the arginine is stabilized by a hydrogen bond between the  $\eta^2$ -nitrogen of the guanidinium group and the backbone carbonyl oxygen of Arg195 (dashed line in fig. 5.4A). In this state the ar/R region is open and allows

5. Is aquaporin-1 a voltage-regulated water channel? \_\_\_\_\_



**Figure 5.4.:** Aromatic arginine (*ar/R*) site of aquaporin-1 with the conserved Arg195 in the up state (A) and in the down state (B). The up state is stabilized by a hydrogen bond between  $\eta^2$ -nitrogen of the guanidinium group and the backbone carbonyl oxygen of Arg195 as well as by a hydrogen bond to the backbone oxygen of Gly125 (dashed line). In the up state a continuous water file is present the *ar/R* region allowing a rapid water flux. In the down state the water column is broken and the permeation of water requires a lever-like motion of the Arg195 side chain (as indicated by the black arrow), rendering the water flux significantly reduced as compared to the up state

the formation of a continuous water file, followed by a rapid water flux. We performed a separate hAQP1 simulation in which the stabilizing hydrogen bond was restrained with an artificial potential. In this simulation, the  $p_f$  was determined to  $7.3 \times 10^{-14} \text{ cm}^3/\text{s}$ . This number is in quantitative agreement to a previous MD simulations using the CHARMM27 force field and the TIP3P water model during which the stabilizing hydrogen bond was also restrained (200).

In the *down* state of Arg195 the water column is broken (fig. 5.4B). Here, a permeation event of a water molecule past the Arg195 requires a lever-like motion of the Arg195 as indicated by the black arrow in fig. 5.4B, which results in a water flux that is significantly reduced as compared to the *up* state. In some simulations the Arg195 even formed an hydrogen bond to the opposing His180 for few nanoseconds, rendering the channel completely closed and eliminating any water flux through the *ar/R* region.

### Membrane potential and Arg195 dihedral state populations

Visual inspection of the simulation trajectories showed that the electrostatic membrane potential  $\Delta\Phi$  has significant influence on the population of the two Arg195 dihedral states. Although both states were present in all simulations, a negative  $\Delta\Phi$  increased the relative population of the *up* state as compared to *down* state,



which correlates with an increased water flux (fig. 5.3).

In a first step, the shift in the relative population can be understood from the fact that the transition between the up and the down state is associated to a short motion of the positively charged guanidinium group in the  $z$ -direction (perpendicular to the membrane). In the down state, the guanidinium group is closer to the cytoplasmic end of the channel (negative  $z$ -direction, compare fig. 5.4). Hence, an electric field pointing in positive  $z$  direction favors the up state. In addition, such an electric field may enhance the lever-like motion of the Arg195 side chain which is required for a permeation event in the down state.

A quantitative estimation for this effect indicates that this direct effect of the electric field on the guanidinium group only partially induces the shift in the population. The shift along the  $z$  coordinate of the guanidinium group is approximately  $\delta z = 2 \text{ \AA}$ . Assuming a membrane potential of  $\Delta\Phi = 200 \text{ mV}$ , a membrane width of  $d = 4 \text{ nm}$ , and a homogenous electric field in the aquaporin, the potential energy difference between the up and the down state is given by  $\Delta V = e \delta z \Delta\Phi / d = 0.4 kT$ . This translates into the relative population by a factor of only  $\sim 0.7$ . Although the electric field might be larger inside the protein since it is more polarizable than the lipid bilayer, the small effect of the electric field on the guanidinium group might not be sufficient to explain the shift in the water flux. Moreover, one would expect an increasing effect on  $p_f$  with increasing  $\Delta\Phi$ , which is not found (compare fig. 5.3). Therefore, more indirect effects may influence the stability of Arg195 in the *up* state. A candidate for such an effect are configurational changes in Loop-C since Arg195 is stabilized in the *up* state by a hydrogen bond to the backbone oxygen of Gly125 in Loop-C (fig. 5.4A). In simulations with large membrane potentials the (negatively charged) Asp127 and Asp130 were pulled into the periplasmic bulk water, inducing a motion of the complete Loop-C. Hence, electrostatic interactions of Loop-C with a positive membrane potential may destabilize the hydrogen bond between Arg195 and Gly125, indirectly increasing the population of the *down* state.

### 5.3. Conclusion

We found strong indications that the water flux through AQP1 can be regulated by an electrostatic membrane potential. The population of the two dihedral states of the conserved Arg195 is a promising mechanism which regulates the water flux.

5. Is aquaporin-1 a voltage-regulated water channel? \_\_\_\_\_

The shift in the relative population between the *up* and *down* states of Arg195 may be induced by the electric force acting on the charged guanidinium group, or by indirect effects such as a destabilization of the Arg195–Gly125 hydrogen bond. An experimental verification of the voltage-regulated water flux through AQP1 is, of course, highly desired.



## 6. Inhibition of aquaporin-1

Malfunctions of aquaporins (AQPs) are involved in a number of pathological conditions such as renal disorders, brain edema, glaucoma, and tumor growth. Numerous studies indicated that upregulation, downregulation, or inhibition of AQPs may be relevant for the treatment of such disorders, making AQPs an interesting target for new drugs. The following section summarizes the role of AQPs in several diseases, particularly highlighting potential medical treatments by AQP inhibition. For more detailed discussions we refer to the reviews by Agre, Castle, and Wang (56, 57, 58).

### 6.1. Introduction — Aquaporins as novel drug targets

**Renal disorders** Seven AQPs facilitate the fluid transport in the kidney (201). In particular, AQP1–3 play an essential role in the concentration of the urine. Humans lacking AQP1 show only a slightly reduced ability to concentrate the urine in case of drought (202). In contrast, mutations in AQP2 can cause a severe disorder called *diabetes insipidus renalis* which is characterized by the excretion of large amounts of diluted urine (48, 203). In mice, the deficiency of AQP3 seemed to downregulate AQP2 expression causing diabetes insipidus (204). Such findings suggest that AQP inhibitors may prove new effective diuretic drugs with less adverse effects than existing reagents.

**Brain edema** Brain injury, stroke, tumors, or hydrocephalus may lead to pathological accumulation of water in the brain which, in turn, may be lethal or cause permanent damage to the brain. Such imbalance in brain homeostasis is likely to be caused by misregulation of AQP4 which is highly expressed in tissues of the blood-brain and brain-cerebrospinal fluid interface (52, 205, 206). AQP4 null mice showed an increased protection against brain edema caused by water intoxication

or reduced blood supply (ischemia) due to cerebral injury (53). On the other hand, recent studies showed that AQP4 plays an important role in the clearance of a different type of—the so called vasogenic— brain edema, which can be caused by infections or brain tumors (52). Mice lacking AQP4 showed a reduced ability to clear a vasogenic brain edema leading to more severe conditions (54). Therefore, both inhibition of AQP4 and upregulation of AQP4 expression would be highly relevant for medical treatment.

**Glaucoma** In the eye, a high intraocular pressure can damage the optic nerve, cause a partial loss of the visual field, or even result in blindness. Worldwide, glaucoma is the second most frequent cause of blindness. The main target in its medical treatment is to reduce the intraocular pressure by reducing the amount of humor that is released into the eye (207). Active transport of solutes across the epithelia creates an osmotic gradient causing diffusion of water into the eye. The fact that AQP1 and AQP4 are highly expressed in the eye's epithelia suggest that the water transport across the epithelium may be facilitated by AQP1 and AQP4. And indeed, AQP1 knockout and AQP1/AQP4 double knockout mice showed a reduced intraocular pressure compared to the wild-type mice (47). Hence, inhibition of AQP1 is expected to reduce the influx of aqueous humor into the eye reducing the intraocular pressure, and is therefore a promising strategy in glaucoma treatment.

**Tumor growth** AQP1 is expressed in the endothelial cells of capillaries in several tumors, suggesting that AQP1 may be involved in angiogenesis (capillary growth) of tumors (49, 55). Experiments on transgenic mice showed that angiogenesis and tumor growth are indeed impaired by the knockout of AQP1 (50, 51). AQP1-deficient mice showed improved survival, indicating that AQP1 may be a novel target for antiangiogenic therapy.

**Exocrine disorders and obesity** In addition, exocrine disorders are related to malfunction of AQPs. For example, abnormal distribution of AQP5 in salivary glands leads to Sjogren's syndrome which is characterized by dry eyes and a dry mouth (208, 209). Moreover, aquaglyceroporins may be related to disorders in the metabolism such as obesity. AQP7 deficient mice showed greater amount of fat mass compared to wild type mice (210, 211), rendering the upregulation of AQP7 a putative strategy for a medical treatment of obesity.



## 6.2. Inhibition of aquaporin-1 by tetraethylammonium

As demonstrated in the previous section, reversible and specific AQP inhibitors are highly desired but not available so far. Mercury, gold, and silver ions are known to inhibit AQP1 by binding to Cys189 in the extracellular pore entrance (212, 213). However, these ions are toxic since they do not bind specifically to AQP1.

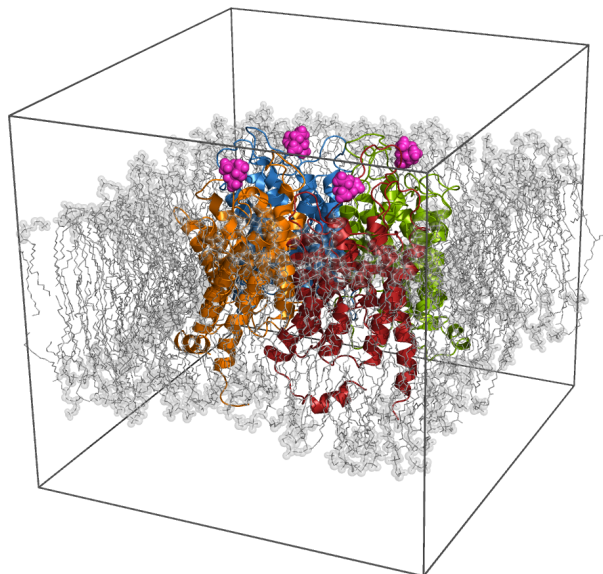
More recently, tetraethylammonium (TEA,  $N(C_2H_5)_4^+$ ) was reported to inhibit AQP1, although its potency to inhibit AQP1 was controversially discussed (214, 215, 216, 217). In oocyte experiments, TEA was reported to reduce the AQP1 water permeability by 20-40% (214). The same group observed that the water permeability of AQP1 expressing kidney-derived cells was reduced by 34% or 50%, respectively, depending on the cell type (217). Detmers *et al.* reported a 44% inhibition of water permeation by TEA with an  $IC_{50}$  concentration of  $1.4 \pm 0.8 \mu M$  (215). In the same study, a putative binding site has been proposed by a combination of docking and MD simulations. Yang *et al.* did not observe any AQP1 inhibition by TEA (216).

Here, we study the effect of TEA on water permeation through human AQP1 to address whether TEA is a possible lead compound for the inhibition of TEA. Using a combination of umbrella sampling simulations, docking, and free MD simulations, we determine the binding site for TEA and estimate the corresponding binding affinity. Noteworthy, binding of a ligand does not necessarily lead to complete inhibition the channel. Therefore, in a second step, the inhibitory effect of TEA on the water flux is determined and compared to the experimental values.

### 6.2.1. Methods

#### Umbrella sampling

To quantify the TEA binding affinity, umbrella sampling simulations were set up as follows. The centers of the umbrella potentials were chosen in steps of  $0.25 \text{ \AA}$  perpendicular to the membrane plane, ranging from the AQP pore entrance into the bulk water. For each umbrella center, a random snapshot was chosen from a 20-ns equilibrium simulation of hAQP1 and used as starting configuration. Within each of the four monomeric channels, a TEA molecule was placed at the umbrella center.



**Figure 6.1:** Typical starting configuration for umbrella simulations. Tetramer of human aquaporin-1 (red, orange, blue, green), membrane of POPE (grey), and TEA (magenta). For clarity, water is not shown.

Water molecules that overlapped with TEA were removed. The simulation box contained the hAQP1 tetramer and four TEA molecules solvated in a membrane of 271 POPE lipids and approx. 19300 TIP4P water molecules (168). The force field and simulation parameters and chosen as described in section 4.2. A typical simulation box is shown in figure 6.1.

Umbrella simulations were carried out by restraining the nitrogen atom of TEA with a harmonic umbrella potential. A force constant of  $1000 \text{ kJ}/(\text{mol nm}^2)$  was applied. In addition, the TEA nitrogen atoms were confined into a cylinder (radius  $r_c = 11 \text{ \AA}$ ) whose axis was centered along the corresponding monomeric pore. The cylindrical confinement was implemented via an additional force  $F(r) = -k_c(r - r_c) H(r - r_c)$  pointing towards the cylinder axis. Here,  $r$  denotes the distance of the TEA nitrogen from the cylinder axis,  $k_c = 400 \text{ kJ}/(\text{mol nm}^2)$  the force constant, and  $H$  the Heaviside step function. Umbrella simulations with the TEA close to the binding region (30 umbrella sections) were run for 6 ns each, simulations with the TEA in bulk water or in loose contact with AQP loops were run for 600 ps each.

Umbrella histograms collected from the simulations were corrected in case of fluctuations of the corresponding monomer within the AQP tetramer. As reference, the z-coordinate (perpendicular to the membrane) of the center of mass of the backbone atoms of the Asn-Pro-Ala (NPA) motifs was applied and defined as  $z = 0$  in figure 6.4. For each of the pores, a free energy profile was computed separately using the weighted histogram analysis method (147). Subsequently, the probability for a



TEA to be located at some position  $z$  along the reaction coordinate was averaged over the four monomers. To this aim, the four profiles  $G_i(z)$  were combined into an effective profile  $G_{\text{eff}}(z)$  via  $\exp(-G_{\text{eff}}(z)/k_B T) = (1/4) \sum_{i=1}^4 \exp(-G_i(z)/k_B T)$ . Here,  $T$  denotes the temperature and  $k_B$  the Boltzmann constant.

### Location of the binding site

In addition, the location of the TEA binding site was determined from the umbrella simulations. The location of the TEA nitrogen was extracted from three umbrella simulations with the TEA located around the minimum in  $G_{\text{eff}}(z)$ . The last 2 ns of these three 6-ns simulations were analyzed. The TEA positions at the four monomers were combined into an effective binding site. The number of TEA positions at a monomer  $i$  that contribute to the effective binding site was chosen proportional to the probability  $\exp(-G_i(z)/k_B T)$  for the TEA to be located at the position  $z$  in monomer  $i$ .

### IC<sub>50</sub> estimates

From the effective profile  $G_{\text{eff}}(z)$  the IC<sub>50</sub> concentration was calculated, i.e. the TEA concentration that leads to a TEA binding with a probability of 50%. To this end, the TEA was considered as bound if the distance  $z$  between the NPA site and the TEA nitrogen was smaller than  $z_b = 25 \text{ \AA}$ . Since the TEA nitrogen was confined into a cylinder of radius  $r_c$  during the umbrella simulations, the IC<sub>50</sub> estimation reduces to the determination of the length  $L$  (or volume  $\pi r_c^2 L$ ) of the cylinder that leads a 50% binding probability.

The probability for the TEA for being bound is given by

$$P_b = N^{-1} \int_{z < z_b} dz \exp(-G_{\text{eff}}(z)/k_B T). \quad (6.1)$$

The probability for being unbound is

$$P_{\text{ub}} = N^{-1} \int_{z_b \leq z < L} dz \exp(-G_{\text{eff}}(z)/k_B T). \quad (6.2)$$

Here,  $N$  is a normalization constant. Applying  $P_b = P_{\text{ub}}$  allows to solve for  $L$  and to determine the corresponding TEA concentration  $\text{IC}_{50} = 1/(\pi r_c^2 L)$ .

### Permeability estimates

To quantify the inhibitory effect of TEA, an equilibrium simulation of hAQP1 with TEA bound into the binding site was compared to a reference simulation without any TEA present. The same simulation system, force field, and simulation parameters as for the umbrella simulations were chosen (except that no umbrella potential and cylindrical confinement was applied). The simulations with and without TEA were run for 10 and 20 ns, respectively. As a measure for water permeability, we counted the number of water molecules that permeated completely across a layer of thickness 9.5 Å including the Asn-Pro-Ala (NPA) site and the aromatic/arginine constriction region. This pore section displays the lowest water diffusion constant (data not shown) limiting the water flux. Hence, the number of complete permeation events across this section is directly proportional to the permeability  $p_f$  (190).

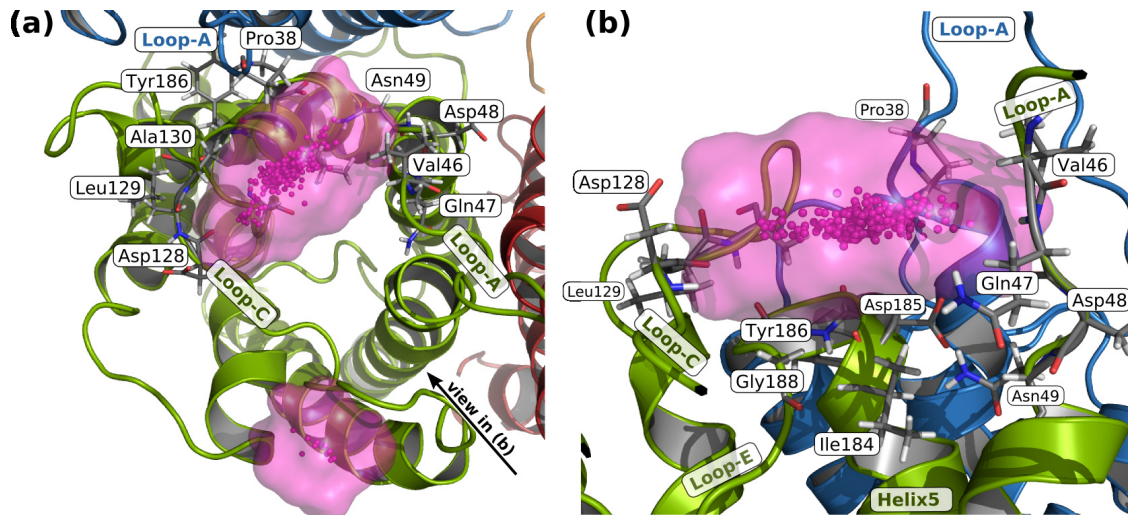
#### 6.2.2. Binding site

The location of the binding site is a prerequisite for the computational design of AQP1 specific inhibitors. Earlier studies reported that the inhibitory effect of TEA on AQP1 was knocked out by mutating Tyr186 into Phenylalanine, and therefore highlighted the role of the E loop in TEA binding (214). Using a combination of docking and MD simulations, a more recent study identified residues in loop A, C and E which may be involved in the binding of TEA. In addition, the A loop of the neighboring monomer was proposed to serve a lid which keeps the TEA inside the binding pocket (215).

Here, we identify the binding site using a combination of three independent methods, i.e, (i) umbrella sampling, (ii) a combination of docking and MD, and (iii) free MD simulations.

The binding site as derived by umbrella simulations is shown in figure 6.2. Observed positions of the TEA nitrogen in the binding site are represented by magenta spheres. The volume which can be occupied by the TEA molecule inside the binding site is indicated as a magenta surface. According to the simulations the TEA is not bound to a single site but samples an extended region in the pore entrance which is located between

- Val46, Gln47, Asp48, and Asn49 of loop A,
- Asp128, Leu129, and Ala130 of loop C,

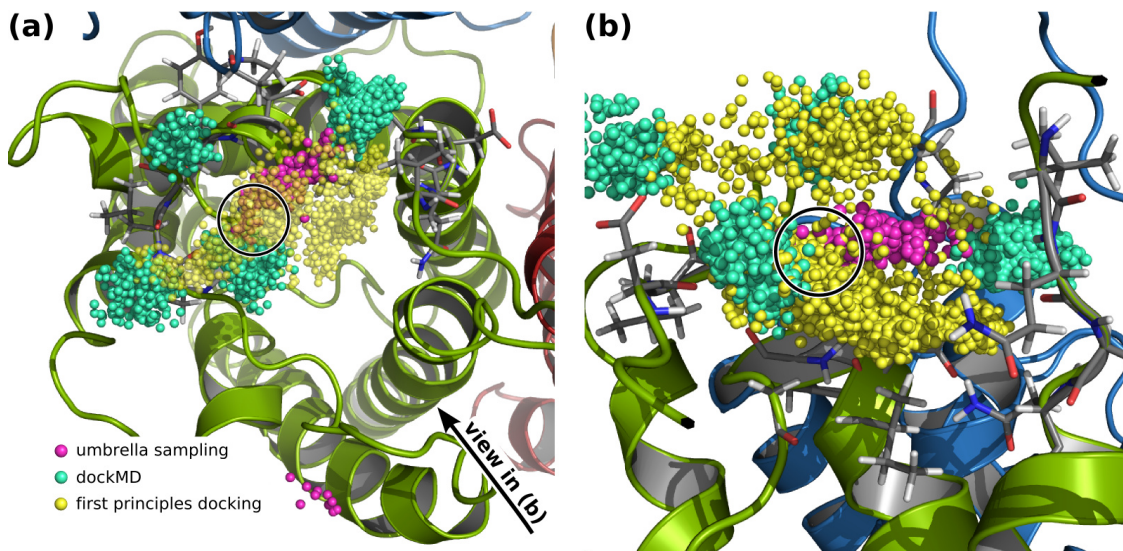


**Figure 6.2.:** Binding site of TEA as derived from umbrella simulations in top view (a) and side view (b). The view in (b) is indicated by the black arrow in (a). Positions of the TEA nitrogen inside the minimum of the potential of mean force (compare figure 6.4) are shown as magenta spheres. The corresponding volume of the complete TEA molecule is shown as a magenta surface. Residues in close contact to TEA are labeled and displayed as sticks. The binding site is not represented by a narrow binding pocket but by a rather extended cloud, even including very few TEA locations outside the channel pore (cloud at the bottom of fig. (a)).

- Ile184, Asp185, and Gly188 of loop E, and
- Pro38 of loop A of the neighboring monomer.

To validate the putative binding site, the binding site derived from umbrella simulations can be compared to TEA locations in free MD simulations, i.e., simulations without any confining potential acting on the TEA molecules. In collaboration with Matthias Müller, two such simulations have been performed which differ in the choice of the initial positions for TEA. For the first simulation, the TEA molecules were placed at favorable positions which have been derived by a docking approach (denoted *dockMD*). TEA nitrogen positions close to the hAQP1 pore are shown as cyan spheres in figure 6.3. For the second set of simulations, 20 TEA molecules were placed at random positions in the bulk water (denoted *first principles docking*). The TEA positions in the binding site as extracted from the latter simulation are displayed as yellow spheres in figure 6.3. The simulation setup for the *dockMD* and first principles docking MD is described in detail in ref. (218).

At first glance, the three approaches yield rather distinct clouds indicating that

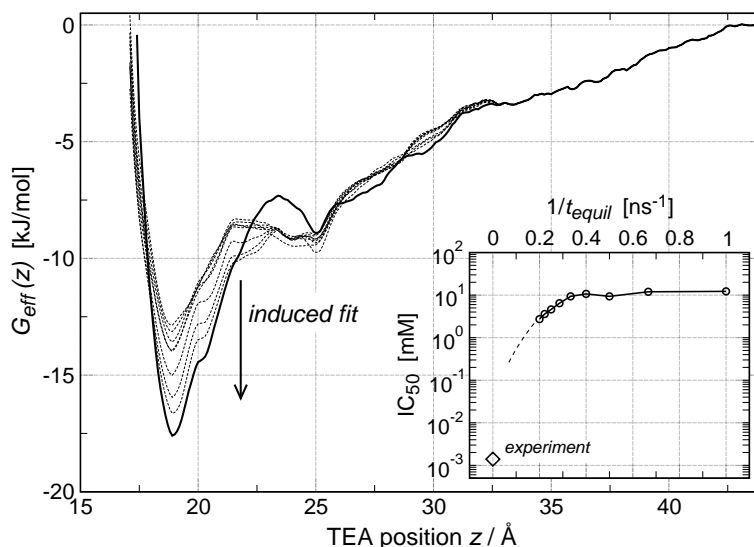


**Figure 6.3.:** Possible binding positions of TEA as derived by three different methods in (a) top view and (b) side view. Umbrella sampling (magenta spheres), unbiased MD simulation with the initial TEA positions derived by docking (denoted ‘*dockMD*’, cyan spheres), and unbiased MD simulation with random initial TEA positions (denoted ‘*first principles docking*’, yellow spheres). The region where the TEA nitrogen positions derived from all three approaches overlap is marked by a black circle. The view in (b) is indicated by the black arrow in (a).

the simulations are not able to produce converged TEA positions within the simulation time. In particular the *dockMD* approach (cyan spheres) sampled TEA positions outside the pore in more or less loose contact with protein loops. Being far away from the narrow part of the AQP pore, such positions are rather unlikely to inhibit the channel. They may, however, represent metastable TEA positions that TEA visits prior to an inhibitory binding event. Likewise, during the *first principles docking* the TEA frequently occupied positions outside the pore (upper yellow cloud in fig. 6.3b) which also likely to represent a preliminary state of binding.

Further inside the channel, the TEA clouds derived from the three approaches significantly overlap as indicated by a black circle in fig. 6.3. This location may therefore be considered as the most likely binding position for TEA. Section 6.2.4 describes the inhibitory effect on the water flux for a TEA molecule located at this position.





**Figure 6.4.:** Potential of mean force (PMF)  $G_{\text{eff}}(z)$  for TEA moving into the binding pore of human AQP1 (black line). The binding site is visible as the minimum in the PMF around  $z \sim 18 \text{ \AA}$ . The pore rearranges upon the presence of a TEA which becomes visible by increasing the time for equilibration  $t_{\text{equil}}$  before analyzing the umbrella simulations (dashed lines). Longer  $t_{\text{equil}}$  only leads to an increasingly pronounced minimum at the binding site but does not change the location of the minimum. The inset displays  $IC_{50}$  concentrations as derived from the PMFs as a function of the inverse  $t_{\text{equil}}$ . The experimental value is shown in diamond representation (215).

### 6.2.3. Binding affinity and $IC_{50}$ estimates

Figure 6.4 displays the potential of mean force (PMF) for TEA moving into the pore of hAQP1, as derived from umbrella sampling simulations. The solid curve shows the PMF that was derived by sampling each umbrella window for 1 ns after an equilibration of 5 ns. The PMF displays a clear minimum at  $z \sim 18 \text{ \AA}$  which corresponds to the binding site displayed in figure 6.2.

Conformational motions of loops occur typically on timescales of tens of nanoseconds and may therefore be insufficiently sampled in nanoseconds umbrella simulations. In particular, the A-loop conformation of the bound state might not be converged, and therefore the actual binding affinity might be larger. To investigate this effect, we calculated the PMF from the 6 ns of simulation with an increasing time for equilibration before analysis. Between 1 ns and 5 ns in steps of 0.5 ns were removed at the beginning of the umbrella trajectories, while the remaining parts of the trajectories were used for the computation of the ‘non-equilibrium’ PMFs (dashed lines in figure 6.4). The position of the minimum does not change

with equilibration time and therefore reflects the (converged) location of the binding site. The binding affinity is observed to increase with simulation length, with an increasingly pronounced minimum in the PMF around  $z \sim 18 \text{ \AA}$ . These findings demonstrate that the TEA binding site rearranges with time in the presence of TEA, displaying a clear example of induced-fit binding at the atomic level.

The inset of figure 6.4 quantifies the induced fit. The graph displays  $IC_{50}$  concentrations, i.e., TEA concentrations at which 50% of the hAQP1 monomers bind a TEA molecule. The  $IC_{50}$  values are shown as a function of the inverse equilibration time  $1/t_{\text{equil}}$  that was used for the calculation of the respective PMF. These  $IC_{50}$  concentrations are rather large, in the range of several millimolar. In the experiments by Detmers *et al.* (215) the  $IC_{50}$  was measured as  $1.4 \mu M$  (diamond). As expected from the PMFs, the  $IC_{50}$  values continuously decrease with increasing equilibration time. Moreover, the plot suggests that the computed  $IC_{50}$  value is likely to decrease further with longer equilibration, as indicated by the dashed line. In this sense, the simulations reveal an upper limit of the  $IC_{50}$  value of  $\sim 3 \text{ mM}$ , which is compatible to, but does not yet quantitatively reflect the experimentally derived value.

#### 6.2.4. Inhibitory effect of TEA on hAQP1

Now that the TEA binding site is reliably identified, we will subsequently address the inhibitory effect on permeation by TEA binding to this site. As described earlier, the binding of TEA at the pore entrance of hAQP1 does not necessarily imply the inhibition of water permeation through the channel. To investigate the inhibition, we analyzed the influence of TEA binding on the water permeability of hAQP1. As a robust measure for permeability, the number of water molecules permeating between the NPA site and the Arg195 was counted, i.e., the permeation events across the narrowest single-file part of the pore (see Methods Section). This number is directly proportional to the osmotic and diffusive permeation coefficients  $p_f$  and  $p_d$  (190).

In a simulation with TEA present in the binding site of hAQP1 we observed only 9 permeation events within 9.5 ns of simulation corresponding to 0.9 events per nanosecond. In a 19 ns reference simulation of hAQP1 without TEA we observed a total of 36 permeation events, corresponding to 1.9 permeation events per nanosecond. Hence, TEA reduced the water flux by  $\sim 50\%$ , which, given the number of



---

events, is significant. This result agrees with oocyte experiments (215).

### **6.3. Summary**

We have located the TEA binding site by a combination of three computational methods. TEA does not bind in terms of a “key and lock” model (219), but rather by an induced fit (220) of the protein to a TEA molecule which is present in the binding site. We showed that a TEA molecule bound to hAQP1 partly inhibits hAQP1, reducing the permeability of hAQP1 by  $\sim 50\%$ . These results confirm TEA as a possible lead compound for hAQP1 inhibition.



# 7. Short range order and collective dynamics of lipid bilayers

## 7.1. Motivation

Lipid membranes in the fluid  $L_\alpha$  phase have since long been studied by elastic x-ray and inelastic neutron scattering as model systems for the more complex and multi-component biological membranes (221, 222, 223, 224, 225, 226, 227, 228, 229). Quantitative information on molecular packing, self-assembly and thermal fluctuations can be obtained (230). Recent studies on such model membranes have been extended to address the collective dynamics by using inelastic x-ray (231, 232) and neutron scattering (233), providing dispersion relations for propagating density modes. Together, elastic and inelastic scattering lead to a quite detailed and quantitative description of lipid bilayers in the reciprocal space. Parallel to experimental techniques, molecular dynamics (MD) simulations have been used to study the structure and dynamics of lipid bilayers (234, 235). MD simulations are routinely used to study the dynamics and energetics of phase behavior, permeation, undulation, membrane self-assembly, and collective dynamics at a molecular level (181, 236, 237, 238, 239, 240, 241).

In the present section we will show that the simultaneous use of MD and diffraction data can be used to refine the structural knowledge of lipid bilayer systems. Simulations can help to interpret diffraction signals in more detail and accuracy than was possible before. At the same time, starting parameters and interatomic potentials used in MD can be verified and refined by comparison with diffraction data.

Elastic x-ray scattering experiments of lipid bilayers usually yield the modulus square of the Fourier transform of the electron density, and thus yield spatial correlations in the electron density rather than the density itself. This fact is well

known as the phase problem in crystallography. In some cases, e.g. structurally simple liquids, the correlation functions themselves — like e.g. structure factors — are the main quantity of interest, from which the nearest neighbor distance and correlation length can be inferred. Already for complex fluids, however, a much more detailed description beyond mere correlation functions is required. Detailed structural information can be obtained by MD simulations. Simulation results can be compared to experiments via a Fourier transformation of the MD coordinates, and can thus facilitate interpretation of the experimental results at a molecular level.

Recently, a number of approaches have been introduced to correlate experimentally measured structure factors (SFs) with SFs extracted from MD simulations to aid molecular interpretation of the experiments. Spaar *et al.* related measured SFs to the interchain distance extracted from MD (242). Additionally, the issue of extracting the area per lipid has been an active field of research (243, 244). Klauda *et al.* (245) addressed this non-trivial issue by relating the area per lipid from simulations to measured density correlations perpendicular to the membrane.

In the present work we combine MD simulations of the well-known phospholipid model system 1,2-dimyristoyl-sn-glycero-3-phosphocholine (DMPC) in the fluid  $L_\alpha$  phase with elastic x-ray and inelastic neutron scattering experiments. We present an analysis of the experimental observables, allowing an extensive comparison of the MD results to a set of scattering data, including x-ray reflectivity, x-ray reciprocal space mappings, and inelastic neutron scattering by triple-axis-spectrometry. Accordingly, details of the vertical density profile  $\rho(z)$  of the acyl chain packing and radial distribution functions, as well as the dispersion relation of the density fluctuations in the plane of the bilayer are addressed. Moreover, the relationship between the experimentally measurable chain correlation peak and the area per head group extracted from MD is investigated. In contrast to Klauda *et al.* (245), who addressed density correlations perpendicular to the membrane, we primarily focus at lateral short-range order of lipids.

In addition to static properties of the bilayer, we present an analysis of collective dynamics of the lipid tails. These have been argued to be crucial for the transport of small molecules through the bilayer (246), and have been investigated experimentally (231, 232, 233) and by MD (240). Generally, collective chain dynamics are interpreted in terms of the three effective eigenmode model, which has, how-



ever, been derived and verified to describe short wavelength density fluctuations of simple liquids like fluid inert gases (247, 248, 249, 250). So far, it is not obvious whether the collective dynamics of lipid tails can be quantitatively described by three effective modes.

The following presentation of our results is organized in three parts:

- (i) First, the electron density profile, the structure factor parallel to the membrane plane, and the reciprocal space mapping of the DMPC bilayer are calculated from equilibrium MD simulation and quantitatively compared to x-ray scattering results.
- (ii) Next, the relationship between the area per lipid and the interchain correlation peak is addressed by employing results derived from simulations at different fixed areas per lipid. Additionally, the MD coordinates are used to verify quantitative relations between the structure factor and the radial distribution function, i.e. between a signal in reciprocal space with structure in real space.
- (iii) Finally, we compare collective chain dynamics in simulation and experiment by analyzing the dynamic structure factor derived from inelastic neutron scattering and MD simulations. The dynamic structure factor is interpreted in terms of the three effective eigenmode model.

## 7.2. Methods

The following section summarizes the methods that were applied to compute structural properties from the MD simulations. The experiments are described in detail in references (233, 251, 252, 253).

**Simulation setup** An equilibrated simulation box of 128 DMPC lipids and 2944 water molecules (fig. 7.1a) was chosen as starting configuration for the simulations. Parameters for DMPC were taken from Berger *et al.* (169) and the Tip4p water model (168) was applied. All simulations were carried out using the Gromacs simulation software (106, 170). Electrostatic interactions were calculated at every step with the particle-mesh Ewald method (114, 115). Short-range repulsive together with attractive dispersion interactions were described by a Lennard-Jones potential, which was cut off at 1.0 nm. The Settle (110) algorithm was used to constrain bond lengths and angles of water molecules, and Lincs (109) was used to constrain the bond lengths of the lipids, allowing a time step of 2 fs. The simulation temperature

was kept constant by weakly ( $\tau=0.1$  ps) coupling the lipids and solvent separately to a temperature bath (116) of 313 K. Likewise, the pressure was kept constant by weakly coupling the system to a pressure bath of 1 bar. In simulations with fixed area per lipid the pressure coupling in  $x$ - $y$  direction (membrane plane) was turned off and the box was coupled to a pressure bath only in  $z$  direction (perpendicular to the membrane). In total, the simulations cover a time of 230 ns.

**Density profiles** The number of water molecules per lipid ( $n_w/n_\ell$ ) to be used in the simulation was determined from the experimental density profile, yielding an estimated ratio  $n_w/n_\ell$  of 27.6. To accomplish this ratio, a simulation box of 128 lipids and 3529 water molecules was set up but used only for comparison of the electron density profiles. The electron density was extracted from a 50 ns trajectory where the first 10 ns were removed for equilibration. The atomic electron densities were approximated by a sum of Gaussians,  $\rho_{el}(\mathbf{r}) = \sum_i a_i^* \exp(-b_i^* \mathbf{r}^2)$ . The parameters  $a_i^*$  and  $b_i^*$  were estimated from the Cromer-Mann parameters  $a_i$  and  $b_i$  (see also next paragraphs) using the fact that a Gaussian form factor in reciprocal space refers to a Gaussian electron density in real space. Fourier transformation yields  $a_i^* = a_i(4\pi/b_i)^{3/2}$  and  $b_i^* = 4\pi^2/b_i$ .

**Structure factors (SF)** In order to achieve a resolution in reciprocal space comparable to the experiments we combined four of the preequilibrated patches of 128 lipids to obtain a simulation system of 512 DMPC lipids and 11776 water molecules ( $n_w/n_\ell = 23$ ). This system was then equilibrated for 12 ns, and simulated for another 18 ns for production. Structure factors were calculated every 10 ps yielding the SF averaged over 1800 frames. The scattering amplitudes at the reciprocal space coordinate  $\mathbf{q}$ ,  $s(\mathbf{q})$ , were calculated by a direct evaluation of the sum  $s(\mathbf{q}) = \sum_n f_n(k) \exp(i\mathbf{q} \cdot \mathbf{r}_n)$ . Here,  $\mathbf{r}_n$  denotes the position of atom  $n$ , and  $f_n(k)$  the non-dispersive part of the atomic form factor of atom  $n$  where  $k = \sin \theta/\lambda$  with the incident angle  $\theta$  and the wavelength  $\lambda$  of the x-ray beam. The  $f_n(k)$  were calculated using the approximation

$$f_n(k) = c^{(n)} + \sum_{i=1}^N a_i^{(n)} \exp\left(-b_i^{(n)} k^2\right), \quad (7.1)$$

where  $c^{(n)}$ ,  $a_i^{(n)}$  and  $b_i^{(n)}$  are the Cromer-Mann parameters which are available in the literature (254). Parameters for  $N = 4$  and  $N = 2$  Gaussians are available.





We used the latter ones. The structure factor is given by  $S(\mathbf{q}) = |s(\mathbf{q})|^2$ .  $S(\mathbf{q})$  was averaged in the  $q_x$ - $q_y$  plane along circles of constant  $q_{\text{lat}} = \sqrt{q_x^2 + q_y^2}$  giving the structure factor  $S(q_{\text{lat}}, q_z)$  as a function of  $q_{\text{lat}}$  and  $q_z$ , corresponding to the components of momentum transfer parallel (in lateral direction) and perpendicular to the membrane, respectively.

The periodic boundary conditions applied in all simulations give rise to an infinite and strictly periodic system in which the resolution in reciprocal space is limited to  $2\pi/L$  where  $L$  is the box size. In our case, the simulation setup of 512 lipids rendered a resolution of  $\sim 0.05 \text{ \AA}^{-1}$  parallel and  $\sim 0.1 \text{ \AA}^{-1}$  perpendicular to the bilayer.

**Radial distribution functions (RDF)** Following the notation in ref. (242) the RDF is defined by

$$g(r) = \frac{n(r, r + \Delta r)}{\rho \Delta V_{2D}}, \quad (7.2)$$

where  $\Delta V_{2D} = 2\pi r \Delta r$  is the volume of a shell with radius  $r$  and thickness  $\Delta r$ , and  $n(r, r + \Delta r)$  is the number of the considered molecules/subgroups in this shell.  $\rho$  denotes their 2D-density and  $r$  is the distance in lateral direction only.

The simulation of 512 DMPC lipids was used to calculate RDFs of (parts of) the head groups and of the lipid tails. Head group RDFs were calculated for each monolayer separately, and subsequently averaged over the two monolayers. For the RDF of the chains the tail region was cut into 60 slices of width  $d_{\text{sl}} = 0.5 \text{ \AA}$ . For each atom in a slice neighboring atoms were counted to contribute to the RDF of the slice if their distance perpendicular to the bilayer was smaller than  $d_{\text{sl}}/2$ . The first maximum in the RDF of the lipid tails  $a$  was extracted after slightly smoothing the RDF by a Savitzky-Golay method. A robust measure of the periodicity of the RDF  $\lambda^{\text{RDF}}$  could be achieved by fitting the maxima of the RDF to quadratic functions.

**Correction in relations between RDF and SF** When calculating properties of the RDF of the lipid tails from their SF a distinction should be made between the RDF of the electron density and the RDF of the centers of the carbon atoms of the lipid tails. The small difference between the two is due to the normalization of the RDF with  $\Delta V_{2D}$  (eq. 7.2): Assuming a carbon atom  $C$  located at a distance  $r_0$  from the origin, with its electron density distributed like a Gaussian with standard deviation  $\sigma$  around  $r_0$ , then the factor of  $1/r$  due to the normalization (eq. 7.2)

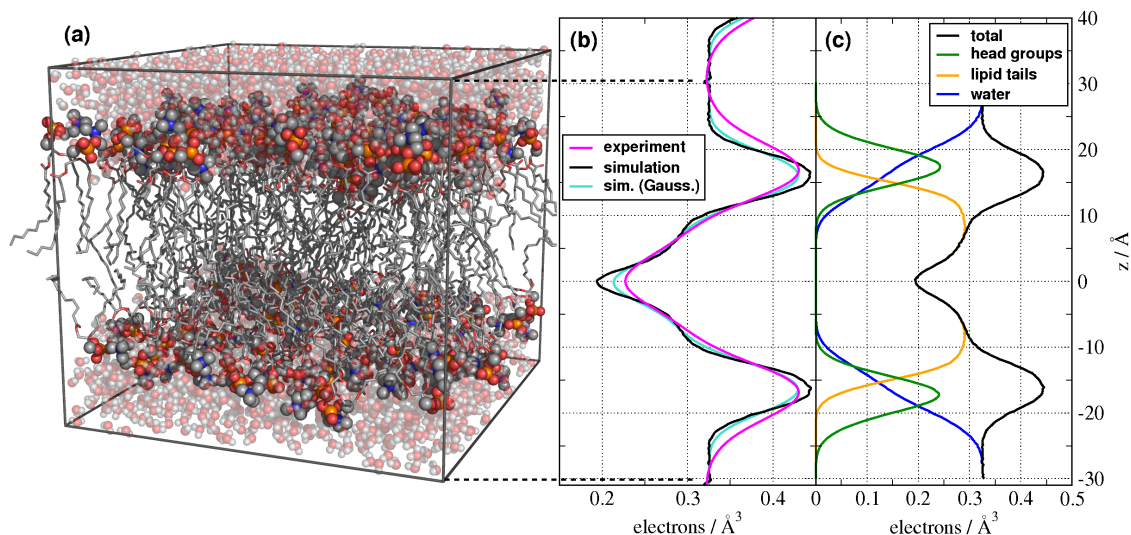
makes the maximum in the RDF of the electron density not be at  $r = r_0$  but at

$$r \simeq r_0(1 - \sigma^2/r_0^2). \quad (7.3)$$

The Cromer-Mann parameters were used to estimate  $\sigma$  in the simulations. A fit of a single Gaussian to the electron density of a united  $\text{CH}_2$  atoms yielded  $\sigma \simeq 0.89 \text{ \AA}$ . As the united  $\text{CH}_2$  atoms are the most frequent atom type in the lipid tails they are the most suitable choice for this estimation. This correction of approx. 3% has been incorporated in fig. 7.6c.

**Statistical errors for the correlation length** All trajectories were subdivided in short trajectories of 1 ns. The SF was calculated for every subtrajectory separately and fitted to a Lorentzian allowing to extract the mean and the standard deviation of the correlation length (assuming a normal distribution).

**Dynamic structure factor (DSF)** A 10 ns simulation of a well equilibrated patch of 128 DMPC lipids and 2944 water molecules was performed to calculate the lateral dynamic structure factor (DSF) of the lipid tails. During the simulation, the nuclear form factor  $F(\mathbf{q}, \tau) = \sum_n f_n \exp(-i\mathbf{q} \cdot \mathbf{r}_n(\tau))$  was calculated every 4 fs. Here,  $f_n$  denotes the nuclear scattering length of atom  $n$ , and  $\mathbf{r}_n(\tau)$  its position at time  $\tau$ . As in the samples used for corresponding experiments (ref. (233)), all  $\text{CH}_2$  and  $\text{CH}_3$  groups of the lipid tails were assumed to be deuterated. The intermediate scattering function was calculated by  $S(\mathbf{q}, t) = N^{-1} \langle F^*(\mathbf{q}, \tau + t) F(\mathbf{q}, \tau) \rangle_\tau$ , where  $N$  denotes the number of atoms, and the brackets the average over the 10 ns simulation, i.e., over  $\sim 2.5$  million pairs of frames, yielding a well converged DSF.  $S(\mathbf{q}, t)$  was calculated up to a time displacement of  $t \leq 100$  ps. Finally,  $S(\mathbf{q}, t)$  was averaged along circles of equal  $q_{\text{lat}}$  (lateral plane) and Fourier transformed yielding the scattering function  $S(q, \omega)$ , and the lateral spectrum  $C_\ell(q, \omega) = (\omega^2/q^2)S(q, \omega)$ . To suppress artifacts from the cutoff at 100 ps,  $S(q, t)$  was turned off smoothly by a Gaussian of width  $\sigma = 20$  ps, which corresponds to a resolution of 0.05 meV. The fit to the three eigenmode model was performed by a least-square Trust-Region algorithm, restricted to  $\omega < 30$  meV.



**Figure 7.1.:** (a) A patch of 128 DMPC lipids and 2944 water molecules was used as starting configuration for the lipid bilayer simulations. To calculate structure factors and radial distribution functions a four times larger patch was simulated. (b) Electron density determined by x-ray diffraction (magenta), by simulation (black), and by convolving the simulation result with a Gaussian (cyan). (c) Electron density extracted from the simulation and decomposed in contributions from head groups (green), lipid tails (yellow), water (blue), and the complete system (black solid).

### 7.3. Electron density

Since this study mainly focuses on the lipid structure parallel to the bilayer, and as the influence of force fields and finite size effects on the electron density have already been investigated in detail (245, 255, 256), we present here only a brief comparison between the electron density perpendicular to the bilayer,  $\rho(z)$ , determined by simulation and x-ray scattering experiments.

Figure 7.1b shows the electron density profile along the membrane normal of a DMPC bilayer. The density from MD simulations (black line) is in good agreement with the density profile from experiment (magenta line). Small differences are observed in form of the more pronounced maxima and minima in the simulation, and a slightly faster decay of the electron density between head groups and bulk water. These differences can be explained either (a) by limitations in the force field, i.e., some aspects of the bilayer structure are not reproduced well, (b) by finite size effects due to the limited size of the simulation box, or (c) by an experimental effect leading to a smearing out of the profile. Smearing out of the experimental

profile may be caused by thermal fluctuations, which are taken into account on the level of a continuum model (257), but residual fluctuation amplitudes on molecular scales may still lead to a flattening of the profile. To test whether smearing can indeed account for the differences between profiles from scattering and simulation, we mimicked smearing by convolving the electron density from simulation with a Gaussian. The best agreement was found by convolving the density from simulation with a Gaussian of width  $\sigma = 1.4 \text{ \AA}$  (fig. 7.1b, cyan line). As expected, the convolution decreases the differences observed at the maxima and minima. The difference in the decay of the density between head groups and water remains. As a next step, we investigated whether the difference in the slope near the head groups is due to finite size effects. A simulation of a bilayer system four times larger than the one shown in fig. 7.1a did, however, not improve the agreement (data not shown). This may be taken as an indication that although the applied force field reproduces the overall measured electron density favorably, not all aspects of the bilayer structure are perfectly reproduced.

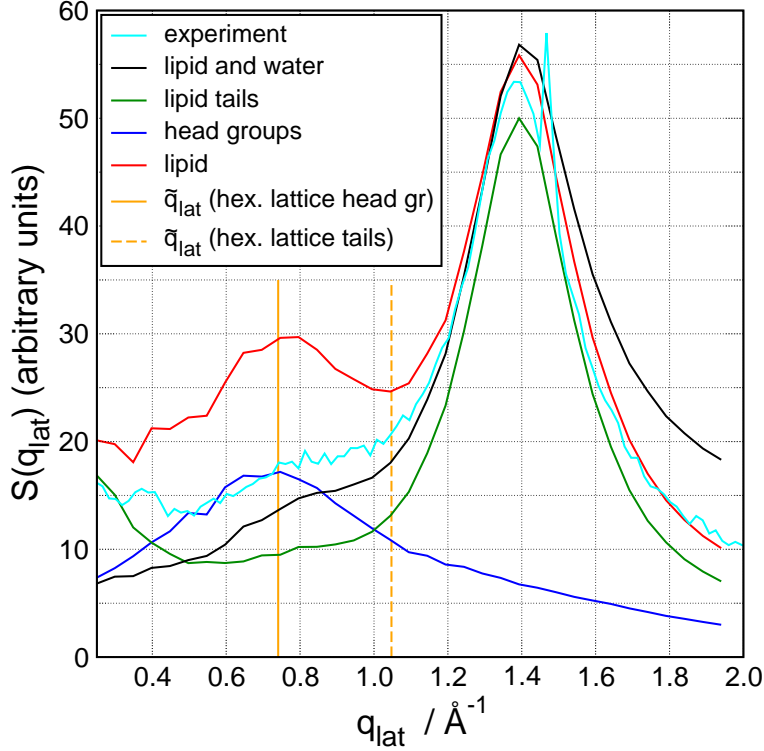
## 7.4. Static structure factors of lipid bilayers

### 7.4.1. Density correlations in the membrane plane

Figure 7.2 shows the static SF  $S(q_{\text{lat}}, q_z = 0)$  representing correlations of the electron density parallel to the membrane. The experimental result from x-ray diffraction (242, 258) is shown in cyan, the simulation result in black. The most pronounced feature of the SF is the chain correlation peak at  $q_{\text{lat}} \simeq 1.39 \text{ \AA}^{-1}$ . Additionally, a shoulder at  $\sim 0.8 \text{ \AA}^{-1}$  is visible.

Taking advantage of the possibility to omit atoms when calculating the SF from the simulation we can isolate the contributions from individual groups to the SF. To this end, we decomposed the SF into contributions from the lipid tails (fig. 7.2, green), the head groups (blue), or the complete DMPC lipids (red). All curves except the head groups show the the peak around  $\sim 1.39 \text{ \AA}^{-1}$ . The observation that the head group curve (blue) does not show this peak, whereas the lipid tails (green) produces this peak only, clearly confirms the common interpretation of this peak as the chain correlation maximum.

In order to quantify the comparison between experiment and simulation, we fitted



**Figure 7.2.:** Static structure factor  $S(q_{\text{lat}}, q_z = 0)$  revealing density correlations parallel to the bilayer. The interchain correlation peak at  $q_{\text{lat}} = 1.39 \text{ \AA}^{-1}$  is clearly visible in both the experimental (cyan line) and simulation results (black line). Additionally, a shoulder at  $q_{\text{lat}} \approx 0.8 \text{ \AA}^{-1}$  that corresponds to the electron density of the head groups is visible in both curves. Properties of the structure factor are easier to identify after decomposing it into contributions from the lipid tails (green), the head groups (blue), or the complete DMPC lipids (red). The solid orange line indicates where the head group peak is expected when assuming that the head groups are arranged on a two dimensional hexagonal lattice. The dashed orange line is the corresponding line for the lipid tails.

the interchain correlation peak in the range of  $1.2$  to  $1.8 \text{ \AA}^{-1}$  to a Lorentzian,

$$S(q) = I \frac{\gamma^2}{(q - q_0)^2 + \gamma^2} + y_0, \quad (7.4)$$

where  $I$  denotes the intensity,  $\gamma$  the half width at half maximum,  $q_0$  the center. Only in simple cases, a peak at  $q_0$  refers to a regular ordering of the density with spacing  $2\pi/q_0$ .  $y_0$  is the baseline of the peak and is generated by, e.g., thermal fluctuations and contributions from inelastic scattering. For a Lorentzian the correlation length is given by  $\xi = 1/\gamma$ . The fit of the chain correlation peak to a Lorentzian (eq. 7.4) was carried out by applying a least squares method using a

Peak	Data	$q_0 / \text{\AA}^{-1}$	$\xi / \text{\AA}$	Color in fig. 7.2
chains	experiment	1.385	5.7	cyan
	simulation (lipid tails)	1.396	5.6	green
	simulation (complete lipids)	1.392	5.1	red
	simulation (lipids & water)	1.411	6.0	black
head groups	simulation (head groups)	0.73	3.2	blue

**Table 7.1.:** Positions of the peak maxima  $q_0$  and the corresponding correlation lengths  $\xi$  of the chain-chain and the head group correlation peak, extracted from x-ray diffraction experiments and MD simulations.

Trust-Region algorithm, yielding the peak position  $q_0$  and the correlation length  $\xi$ . This fit has been performed for (i) the experiment, (ii) the coordinates of the lipid tails, (iii) of the complete lipids, and (iv) of lipid and water. The results are summarized in table 7.1. The peak position as well as the correlation length are in favorable agreement with experiment. Since the level of hydration was slightly lower in experiment compared to simulation we performed an additional simulation with a water to lipid ratio of  $n_w/n_\ell = 15$  which reproduces the experimental bilayer repeat distance of  $50 \text{\AA}$ . The results differ less than 1% compared to the results in table 7.1 (data not shown) showing that in the regime of partial hydration the exact level of hydration has no significant influence on the lipid tail packing.

A second feature in the experimental SF (fig. 7.2, cyan line) and the SF of the simulation system (black line) is the head group peak, which is only visible as a shoulder at  $\sim 0.8 \text{\AA}^{-1}$ . Omitting the water (red line) leads to a significantly more pronounced peak, indicating that gaps between head groups are filled by water, smearing out the electron density correlation. When taking only the DMPC head groups into account (blue line) the head group correlation is visible as a single peak, which can again be fitted to a Lorentzian (eq. 7.4) giving the position of the peak maximum and the correlation length. The fit in the range of  $0.4 \text{\AA}^{-1}$  to  $1.4 \text{\AA}^{-1}$  yields  $q_0 = 0.72 \text{\AA}^{-1}$  and  $\xi = 3.2 \text{\AA}$  (see also table 7.1) showing that although the head group correlation is present, it is weak.

In summary, the simulation result agrees favorably with experiment and confirms the common interpretation of the chain correlation peak and the head group shoulder. Having identified and analyzed the SF the following questions arise: How do these structure factor maxima relate to specific configurations of lipids in real



space? In particular, to which degree is the picture of a simple two dimensional (2D) fluid valid to describe the short range order of the lipid tails or the head groups?

In a first step, we compare the lipid head group distribution to a grid of spheres arranged on a 2D hexagonal lattice. The distance between the centers of adjacent head groups is then  $\tilde{d} = (2A_L/3^{1/2})^{1/2}$  where  $A_L$  is the area per lipid, and a peak at  $\tilde{q}_{\text{lat}} \approx 2\pi/\tilde{d}$  is expected in the SF. In our simulations,  $A_L$  equaled  $62.3 \text{ \AA}^2$  resulting in  $\tilde{q}_{\text{lat}} = 0.74 \text{ \AA}^{-1}$  as indicated by a solid orange line in fig. 7.2.

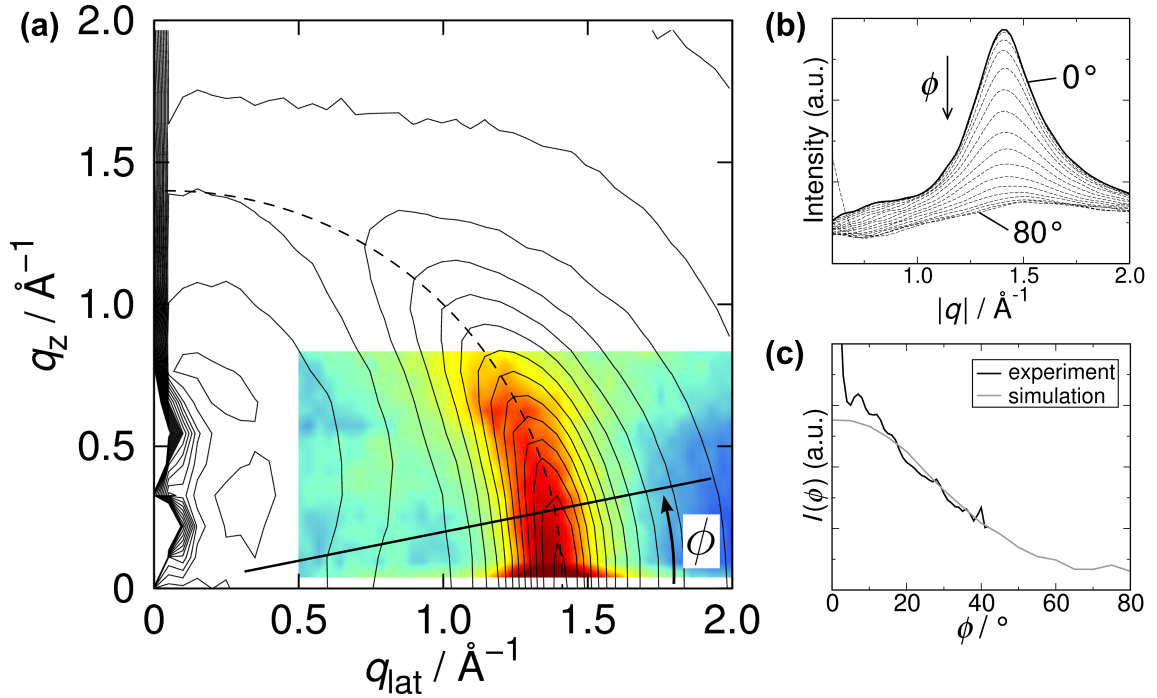
And indeed, the true position of the head group peak at  $0.72 \text{ \AA}^{-1}$  is in good agreement with the position expected from this simplistic hexagonal lattice model.

In contrast, the idea of a densely packed 2D fluid does not hold for the lipid tails. Using the same model the chain correlation peak would be expected at  $1.05 \text{ \AA}^{-1}$  (dashed orange line) which is far off the peak at  $1.39 \text{ \AA}^{-1}$ . Hence, the distance between lipid tails is shorter than expected from the homogeneous hexagonal lattice model. This short distance may be due to formation of local clusters of lipid tails. Apparently, in the fluid  $L_\alpha$  phase, it is not possible to relate the area per lipid straightforward to the interchain distance. It would be interesting to investigate this issue also in the gel phase, where the area per lipid is more reduced, thus shifting the dashed orange line in fig. 7.2 to higher  $\tilde{q}_{\text{lat}}$ .

## 7.4.2. Including correlations perpendicular to the membrane: reciprocal space mapping

So far, we have considered  $S(q_{\text{lat}}, q_z = 0)$ , reflecting the lateral correlations in the plane of the bilayer in a range of  $q_{\text{lat}}$  corresponding to molecular length scales. However,  $S(q_{\text{lat}}, q_z = 0)$  characterizes only the two dimensional order of lipid molecules, after projection of the electron density into the plane of the membrane. A full description of the fluid structure is obtained, by considering the 3-dimensional structure of the tails, including their tilt degree of freedom. To this end,  $S(q_{\text{lat}}, q_z = 0)$  needs to be complemented by a fully two-dimensional reciprocal space mapping (RSM)  $S(q_{\text{lat}}, q_z)$  around the chain correlation peak (242). Experimentally, this is only possible for highly aligned membranes, where powder averaging does not destroy the separation of the different  $q$  components.

The RSM of the DMPC bilayer is displayed in fig. 7.3a. The simulation results are shown as a contour plot. For comparison, the x-ray data are shown as a color



**Figure 7.3.:** (a) Reciprocal space mapping (RSM) of DMPC, i.e. the 2D structure factor  $S(q_{\text{lat}}, q_z)$  representing density correlations parallel and perpendicular to the bilayer: simulation result as contour, x-ray data as color coded plot. (b) Radial slices of the RSM from simulation along angles  $\phi$  as indicated by the line in (a).  $\phi$  is varied from  $\phi = 0^\circ$  (highest maximum, solid line) to  $\phi = 80^\circ$  (lowest maximum). (c) Peak maximum of radial slices versus the angle  $\phi$ , extracted from simulation (shaded) and experiment (black line).

coded plot. Again, favorable agreement is found in the full range of available experimental data. Both the simulation and x-ray data show the extension of the interchain correlation peak along a circle with radius  $\sim 1.4 \text{\AA}^{-1}$  around the origin of reciprocal space, indicated by a dashed line in fig. 7.3a. The bent peak shape originates from lipid tail populations (or segments) which are tilted against the membrane normal.

A more quantitative comparison is possible by analyzing  $S(q_{\text{lat}}, q_z)$  along radial slices in the  $q_{\text{lat}}-q_z$  plane. Figure 7.3b shows the simulation RSM along these slices at angles  $\phi$  (as defined in fig. 7.3a) between  $0^\circ$  and  $80^\circ$  in steps of  $5^\circ$ . The maximum of the chain correlation peak clearly decreases and its half width (HWHM) increases with increasing  $\phi$ . This is an indication that the lateral correlation length of the tail decreases the more the tails are tilted with respect to the membrane normal. The decay of the chain correlation peak along  $\phi$  can be quantified by fitting the





experimental and simulation data to a Lorentzian (eq. 7.4) at various angles  $\phi$ . The resulting peak maxima  $I(\phi)$  are displayed in fig. 7.3c. Simulation and experiment may only be compared in the  $10^\circ \lesssim \phi \lesssim 40^\circ$  range. In this range, good agreement is observed. For  $\phi < 10^\circ$  corresponding to exit angles equal to the critical angle  $\alpha_f = \alpha_c$ , the Fresnel transmission function leads to an enhancement of intensity, which is a well known optical effect in grazing incidence diffraction, rendering higher experimental than theoretical values.

## 7.5. Radial distribution functions

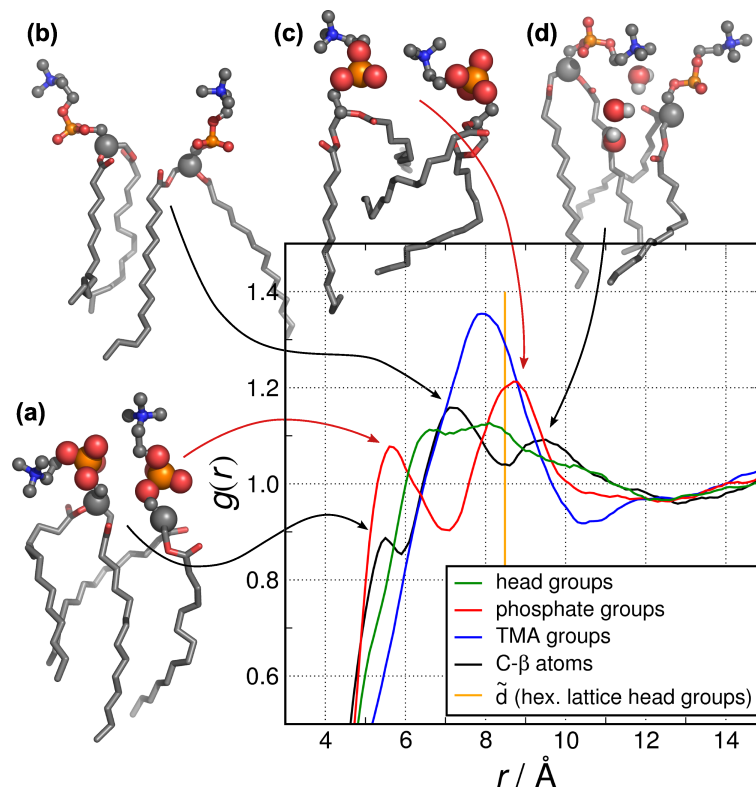
### 7.5.1. Head group structure

The RDF (eq. 7.2) of the lipid head groups yields a more detailed picture of their short range order. Here, we present the RDFs computed from the centers of mass, which is more pronounced and less smeared out than the RDF computed from the electron densities. Thus, it helps to identify features in the structure.

The RDF of the complete head groups (fig. 7.4, green) reveals that there is not one but rather two or more typical distances of adjacent head groups. (Note that ‘distance’ always refers to COM distances). Apparently, the structure of lipid head groups is more complex than that of spheres in a 2D liquid and is not well described by the RDF of the COM of the complete head groups. Therefore, we proceed to calculate the RDF of three subgroups of the head group: the phosphate group  $\text{PO}_4^-$ , the trimethylammonium (TMA) group  $\text{N}(\text{CH}_3)_3^+$ , and the central carbon atom of the glycerol (denoted C- $\beta$ ), i.e, the carbon atom to which the lipid tails are bound. Interestingly, these three subgroups show very different short range correlations.

The simplest short range structure is formed by the TMA groups producing a single peak in the RDF at a lateral distance of  $r \simeq 7.9 \text{ \AA}$  (fig. 7.4, blue). Presumably this is due to the situation that the TMA group at the very top of the head group is least constricted in its motion by interactions with other lipid molecules. In the same fashion as above, we compare our results to the 2D hexagonal lattice model (orange line in fig. 7.4). For the TMA group we find that its peak maximum is only slightly below the lipid distance in the hexagonal lattice model  $\tilde{d}$ , indicating that the TMA groups show a weak tendency arrange in a non-homogenous distribution.

The RDFs of the phosphate (red line) and the C- $\beta$  atoms (black line) show more complex behavior with two and three maxima for adjacent head groups, respec-



**Figure 7.4.:** Radial distribution functions  $g(r)$  of the centers of mass of molecular components of the head groups: complete head groups (green), phosphate (red), trimethylammonium groups (blue), and the central carbon atom of the glycerol (black, denoted by  $C-\beta$ ). The radial distribution functions reveal complex structural determinants for the packing of (components of) the head group. The distance that is expected if the head groups were ordered on a two-dimensional hexagonal lattice is indicated in orange.

tively. This behavior is related to the fact that (a) these atoms are confined by steric interactions with other lipid molecules, and (b) that lipid molecules are not rotationally symmetric.

To illustrate the complex short range order of the head groups we selected representative pairs of lipid molecules from MD snapshots that contribute to the peaks in the RDFs. They are displayed in fig. 7.4 with arrows indicating to which peak they contribute. The first peak in the phosphate RDF at  $r \simeq 5.6 \text{ \AA}$  can be assigned to two directly neighboring phosphates (fig. 7.4a) whereas the second one at  $r \simeq 8.7 \text{ \AA}$  corresponds to two phosphates separated by a TMA group (fig. 7.4c). This TMA can either belong to one of the two lipids or be part of a third lipid. The first maximum in the  $C-\beta$  RDF at  $r \simeq 5.5 \text{ \AA}$  corresponds to pairs of lipids that are tightly packed with their lipid tails relatively parallel (fig. 7.4a). Although the



peak is clearly present, its maximum is quite low indicating that this configuration is rare. The second and most prominent maximum of the C- $\beta$  RDF at  $r \simeq 7.2 \text{ \AA}$  belongs to a broad set of disordered adjacent lipid molecules. Figure 7.4b shows a representative configuration for this group of poorly packed, but very frequent structures. The third peak around  $r \simeq 9.5 \text{ \AA}$  usually corresponds to two lipid molecules being separated by either a few water molecules (fig. 7.4d), or a tail of a third lipid molecule (not shown).

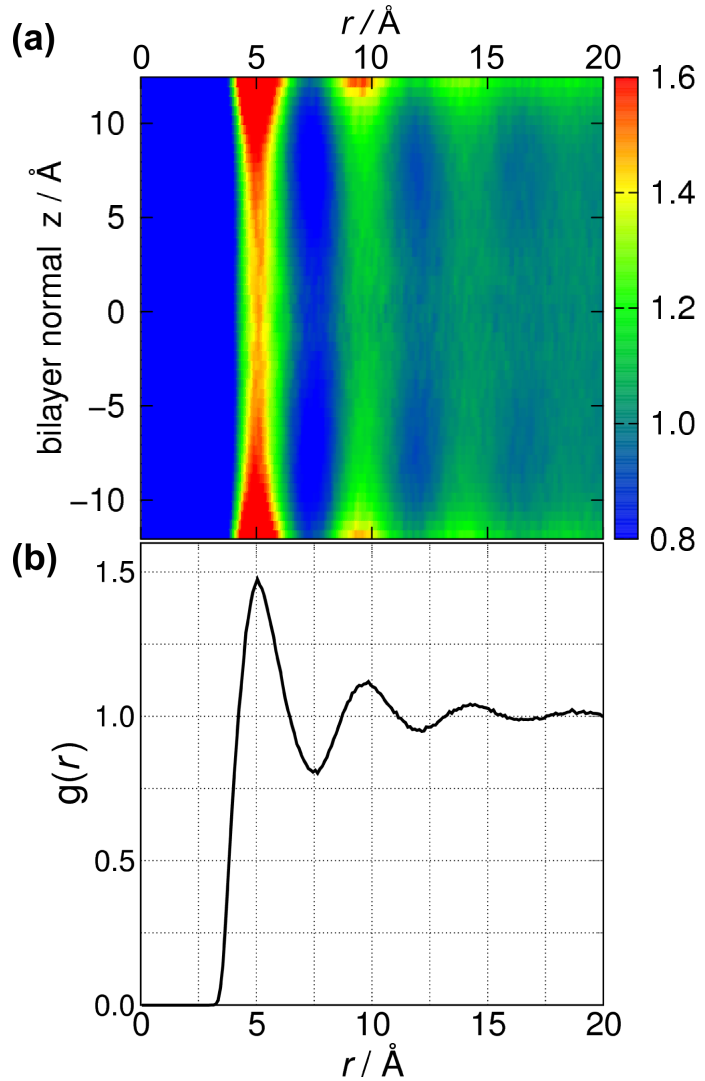
In summary, these results indicate that the picture of head groups (in contrast to the lipid tails structure) being ordered like a dense  $2D$  liquid is a reasonable first order approximation. On a closer look however, packing effects of the asymmetric lipids, and interactions of the head group sub-units render the energy landscape of the lipids much more complex than that of a simple liquid. The impact of this complex energy landscape on dynamic processes like bilayer formation or lipid diffusion remains a topic for future studies.

## 7.5.2. Radial distribution function of lipid tails

Figure 7.5a shows the RDF of the lipid tails in slices parallel to the bilayer extracted from a 20 ns equilibrium simulation of 512 DMPC molecules. For the calculation, the C atoms of the tails have been treated as points, to determine the RDF of the tail center positions. A pair of C atoms contributing to the RDF may be part of the same lipid molecule, but not of the same tail. The results show that the RDF is a robust function for  $-10 \text{ \AA} \lesssim z \lesssim 10 \text{ \AA}$ , i.e., it is almost independent of the distance from the center of the bilayer at  $z = 0$ . This confirms the findings of a previous study (235, 242) on POPC (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine) that was based on only a single MD snapshot. Hence, averaging the RDF around the central region of the bilayer, i.e. for  $|z| < 7.5 \text{ \AA}$  leads to a robust RDF of the lipid tails, as shown in fig. 7.5b.

## 7.6. Lipid tail packing and area per lipid

The main goal of this section is to relate the experimentally measured SF to intrinsic structural quantities of the bilayer, in particular to the area per lipid  $A_L$  and the radial distribution function of the lipid tails. Two problems are addressed: First, we investigate how the chain correlation peak of the SF depends on  $A_L$ , and whether

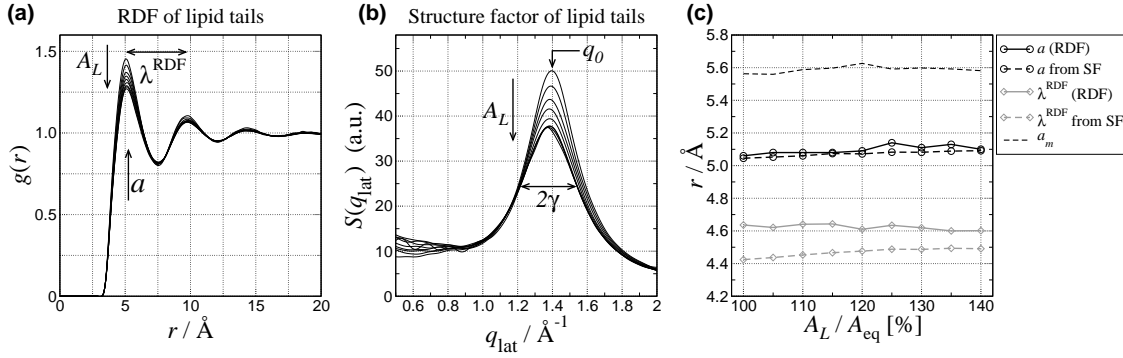


**Figure 7.5:** (a) Radial distribution function  $g(r)$  of the lipid tails in slices parallel to the DMPC bilayer.  $z = 0$  corresponds to the center of the membrane.  $g(r)$  is almost independent of  $z$ . In particular, the positions of the maxima and minima are robust properties of the lipid bilayer. (b) Slice-wise radial distribution function in (a) averaged over  $|z| < 7.5 \text{ \AA}$  around the center of the bilayer.

the latter can be estimated from the peak position and shape. Second, quantitative relations between properties of the SF and RDF shall be tested at varied  $A_L$ .

### 7.6.1. Invariance of RDF and structure factor as a function of the area per lipid

A patch of 512 DMPC and 11776 water molecules was first simulated for 20 ns at constant pressure allowing to extract the equilibrium area per lipid  $A_{\text{eq}} = 62.3 \text{ \AA}^2$  under the simulation conditions. Then, the system was simulated with the area per lipid  $A_L$  being constrained between 100% and 140% of  $A_{\text{eq}}$  with steps of 5%. Each of these 9 systems was simulated for another 20 ns. We restricted to  $A_L \geq A_{\text{eq}}$

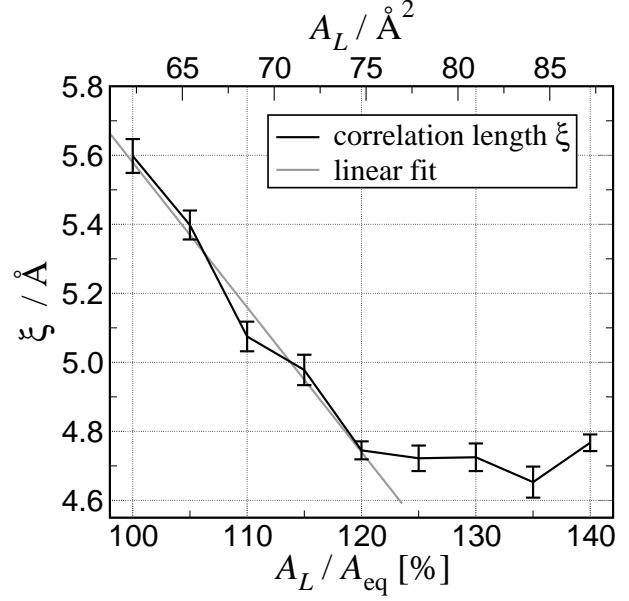


**Figure 7.6.:** (a) Radial distribution function (RDF) and (b) structure factor (SF) of lipid tails from simulations at constant area per lipid  $A_L$ .  $A_L$  was constrained between 100% and 140% of the equilibrium area per lipid and increased in steps of 5% ( $A_{\text{eq}} = 62.3 \text{ \AA}^2$ ). (c) The position of the first maximum  $a$  of the RDF (black solid line, circles),  $a$  calculated from the SF using eq. 7.6 (black dashed line, circles), periodicity  $\lambda^{\text{RDF}}$  of the RDF (shaded solid line, diamonds),  $\lambda^{\text{RDF}}$  calculated from the SF using eq. 7.7 (shaded dashed line, diamonds), and the mean interchain distance  $a_m$  (dashed black line). The quantities are plotted versus the area per lipid in units of  $A_{\text{eq}}$ .

since bilayers of the size of our simulation box were found to form bends within only a few 100 ps when being constrained to  $A_L < A_{\text{eq}}$  rendering the simulations unsuitable for the the following quantitative analysis.

Figure 7.6a shows the RDF  $g(r)$  of the 9 simulations at different areas per lipid. The first maximum in the RDF  $a$  is the most likely interchain distance, which must be distinguished from the mean interchain distance  $a_m = \int_{1\text{sh}} r g(r) dr / \int_{1\text{sh}} g(r) dr$ . Here,  $\int_{1\text{sh}}$  is an integral over the first shell of lipid tails, i.e., up to the first minimum of  $g(r)$ .  $\lambda^{\text{RDF}}$  denotes the periodicity of the RDF as depicted in fig. 7.6a.

Interestingly, only the value of  $g(r)$  at the maxima and minima of the RDF depends on the area per lipid  $A_L$ , but not the position of the maxima and minima or the periodicity  $\lambda^{\text{RDF}}$  of the RDF. This is also shown in fig. 7.6c where the position of the first maximum  $a$  of the RDF is plotted as a black solid line versus the relative increase of the area per lipid showing that  $a$  is independent of  $A_L$ . Similarly, the periodicity  $\lambda^{\text{RDF}}$  (shaded solid line) is independent of  $A_L$ . The same trend can be observed in the structure factor of the lipid tails (fig. 7.6b); with increasing  $A_L$  the chain correlation becomes less pronounced, but the position of the maximum in reciprocal space  $q_0$  is almost constant. Apparently, when increasing  $A_L$ , the lipid bilayer does not undergo a uniform expansion. Instead, lipid tails remain at constant relative positions and only the population of interchain distances is



**Figure 7.7:** Correlation length  $\xi = 1/\gamma$  versus the area per lipid  $A_L$  in units of the equilibrium area per lipid  $A_{\text{eq}}$  (solid line). The error bars display standard deviations. For  $A_L \lesssim 1.20 A_{\text{eq}}$  the membrane stays intact allowing to extrapolate the  $\xi$ - $A_L$  dependence (shaded line).

reduced. In other words, the bilayer responds in a lateral *inhomogeneous* manner to the decrease of pressure. This can be understood from the fact that the interaction between the chains is predominantly attractive.

Although interchain distances are obviously not explicitly dependent on  $A_L$ , we cannot exclude that under environmental conditions (temperature, humidity, salt) that induce a different  $A_L$  the chain correlation might also be affected. From our results we conclude that it is not the area per lipid itself that determines the periodicity of the RDF (or the position of the interchain correlation peak of the SF). This finding underlines the fact that for a given type of lipid (and a given temperature) the  $q_0$ ,  $a$ ,  $a_m$  and  $\lambda^{\text{RDF}}$  are robust properties of the bilayer. It also rules out any possibilities to deduce  $A_L$  experimentally from the measured peak position of the SF.

### 7.6.2. Linear dependence between correlation length and area per lipid

While  $A_L$  does not determine  $a$  and the corresponding  $q_0$ , it does have an effect on the correlation length  $\xi = 1/\gamma$  of the chain-chain ordering as shown in fig. 7.7. When increasing  $A_L$  from  $62.3 \text{\AA}^2$  ( $1A_{\text{eq}}$ ) to  $74.8 \text{\AA}^2$  ( $1.2A_{\text{eq}}$ ),  $\xi$  decreases linearly from  $5.7 \text{\AA}$  to  $4.85 \text{\AA}$ . When pushing the bilayer even further away from equilibrium we reach a non-linear regime where  $\xi$  seems to saturate at  $\sim 4.8 \text{\AA}$ . Visual



inspection of the simulation trajectories showed that under such extreme conditions the bilayer starts to rupture. Lipid molecules form clusters and water molecules penetrate into the membrane between the clusters. For  $A_L \lesssim 1.20 A_{\text{eq}}$  the membrane remains intact and allows to quantify the  $\xi$ - $A_L$  dependence empirically by a simple linear model  $\xi(A_L) = \alpha A_L + \xi_0$ . From the simulation data we estimate  $\xi_0 = 9.78 \text{ \AA} (\pm 0.4\%)$ , and  $\alpha = -0.0674/\text{\AA} (\pm 0.8\%)$  (fig. 7.7, shaded line) by a simple linear least squares fit. (Errors correspond to standard deviations.) This linear  $\xi$ - $A_L$  dependence can be applied to estimate the experimental area per lipid  $A_L^{\text{exp}}$  from the correlation length  $\xi^{\text{exp}}$  determined from the SF. A least squares fit of the experimental SF to a Lorentzian results in  $\xi^{\text{exp}} = 5.74 \text{ \AA} (\pm 1.4\%)$  (see also table 7.1) yielding  $A_L^{\text{exp}} \approx 59.9 \text{ \AA}^2 (\pm 2.4\%)$ . Note that the value of  $A_L^{\text{exp}}$  refers to the experimental conditions (temperature, humidity) and not to our simulation setup. The uncertainty of 2.4% has been determined by usual error propagation and accounts for statistical errors only. Within the reported uncertainty,  $A_L^{\text{exp}}$  agrees with the experimentally determined value of  $59.6 \text{ \AA}^2$  (244) as well as with experimental results that have been refined by simulation methods ( $60.6 \text{ \AA}^2$ , ref. (245)). This analysis shows that the interchain correlation length is an observable that can be quantitatively related to the area per lipid.

### 7.6.3. Determination of structural parameters from the experimental structure factor

Assuming that the SF is a Lorentzian (eq. 7.4), and  $\gamma/q_0 \ll 1$  holds, the RDF in 2D can be well approximated by (242)

$$g(r) \simeq 1 + \frac{I\gamma}{\rho} \sqrt{\frac{q_0}{2\pi}} \frac{e^{-r/\xi}}{\sqrt{r}} \left[ \cos(q_0 r - \pi/4) + \frac{\gamma}{2q_0} \cos(q_0 r + \pi/4) + O\left(\frac{\gamma^2}{q_0^2}\right) \right]. \quad (7.5)$$

In our case, the term of order  $\gamma/q_0$  is a correction of only 7% to the first cosine. From this expression the position of the first maximum of  $g(r)$  is calculated to be at

$$a \simeq \frac{9\pi}{4q_0} - \frac{3\gamma}{2q_0} \quad (7.6)$$

and the period of the RDF is

$$\lambda^{\text{RDF}} = 2\pi/q_0. \quad (7.7)$$

In order to verify these relations, we employ the RDFs and SFs obtained from our simulations at different areas per lipid.  $a$  and  $\lambda^{\text{RDF}}$  as calculated by eqs. 7.6 and 7.7 are compared to the corresponding values obtained directly from the RDF. Figure 7.6c shows these structural parameters as a function of  $A_L$  in units of the average area per lipid in the equilibrium simulation  $A_{\text{eq}}$ . The black solid line (with circles) shows the most likely interchain distance  $a$ . The black dashed line (with circles) depicts  $a$  calculated from the SF of the lipid tails using eqs. 7.6 and 7.3. We find favorable agreement with a deviation of  $\sim 0.5\%$ . In addition, the dashed black line (no symbols) shows the mean interchain distance  $a_m$ , which is found to be approximately 9.5% larger than  $a$  and does also not depend on  $A_L$ . The shaded lines in fig. 7.6c show the periodicity  $\lambda^{\text{RDF}}$  of the RDF as determined either from the local maxima of the RDF itself (shaded solid line), or from the SF of the lipid tails (shaded dashed line). For the latter we corrected eq. 7.7 according to eq. 7.3. Again, good agreement within 4% is found indicating that, indeed, the period of the RDF can be determined to high accuracy from the SF of the lipid tails.

Hence, it is straightforward to calculate the structural properties  $a$ ,  $\lambda^{\text{RDF}}$ , and  $a_m$  from the measured SF. Applying the validated equations 7.6 and 7.7 to the experimental SF (see table 7.1) gives  $a^{\text{exp}} = 5.07 \text{ \AA}$  and  $\lambda^{\text{exp}} \simeq 4.5 \text{ \AA}$ . Using the result that  $a_m$  is 9.5% larger than  $a$  renders the mean interchain distance under the experimental conditions to be  $a_m^{\text{exp}} \simeq 5.6 \text{ \AA}$ .

## 7.7. Dynamic structure factor

### 7.7.1. Definitions and theory

In order to clarify the nomenclature we first review some definitions needed to investigate collective dynamics of bilayers. The *spatial and temporal density correlation function* is given by

$$G(\mathbf{r}, t) = N^{-1} \int d^3\mathbf{r}' \langle \rho(\mathbf{r}', 0) \rho(\mathbf{r}' + \mathbf{r}, t) \rangle, \quad (7.8)$$





where  $N$  is the particle number, and the brackets  $\langle \dots \rangle$  denote the ensemble average. In the case of neutron scattering, the corresponding density is given by

$$\rho(\mathbf{r}, t) = \sum_n f_n \delta(\mathbf{r} - \mathbf{r}_n(t)) \quad (7.9)$$

with the position  $\mathbf{r}_n(t)$  of particle  $n$  at time  $t$ , and  $f_n$  its nuclear scattering length.

The *intermediate scattering function* (in some publications termed density correlation function) is the spatial Fourier transform of  $G(\mathbf{r}, t)$ , i.e.,

$$\begin{aligned} S(\mathbf{q}, t) &= \int d^3\mathbf{r} e^{i\mathbf{q}\cdot\mathbf{r}} G(\mathbf{r}, t) \\ &= N^{-1} \sum_{n,m} f_n f_m \langle \exp\{i\mathbf{q} \cdot [\mathbf{r}_n(t) - \mathbf{r}_m(0)]\} \rangle. \end{aligned} \quad (7.10)$$

The *scattering function* in turn is defined as the temporal Fourier transform of  $S(\mathbf{q}, t)$ , i.e., by

$$S(\mathbf{q}, \omega) = \int_{-\infty}^{\infty} d\omega e^{i\omega t} S(\mathbf{q}, t). \quad (7.11)$$

Note that  $S(\mathbf{q}, \omega)$  is real since  $S(\mathbf{q}, t) = S^*(\mathbf{q}, -t)$  holds for the complex intermediate scattering function.

For a fluid, it has been shown (248) that according to kinetic theory  $S(q, \omega)$  can be decomposed into an infinite set of Lorentzians,

$$S(q, \omega) = \frac{S(q)}{\pi} \sum_{j=-\infty}^{\infty} \text{Re} \left( \frac{A_j(q)}{i\omega + z_j(q)} \right). \quad (7.12)$$

Here,  $S(q)$  is the static structure factor, and the parameters  $A_j$  and  $z_j$  are real or appear in complex conjugate pairs,  $A_{-j} = A_j^*$  and  $z_{-j} = z_j^*$ .  $A_j$  and  $z_j$  are not independent but obey sum rules,

$$\sum_{j=-\infty}^{\infty} A_j(q) [z_j(q)]^n = R_n(q), \quad (7.13)$$

where  $R_0(q) = S(q)$ ,  $R_1(q) = 0$ , and  $R_2(q) = -k_B T q^2 / M$  with the temperature  $T$ , the Boltzmann constant  $k_B$ , and the particle mass  $M$ .

In the hydrodynamic limit, i.e. for wavelengths much larger than the diameter of the particles, only three Lorentzians contribute to  $S(q, \omega)$ , and the parameters  $A_j$  and  $z_j$  can be expressed in terms of hydrodynamic quantities, which are, e.g.,

summarized in ref. (248). Then the model consists of a central Rayleigh line ( $j = 0$ ), and two Brillouin lines ( $j = \pm 1$ ), which correspond to the heat mode and two sound modes, respectively. Therefore, it is usually referred to as the three effective eigenmode model. It has been argued that it is not the particle diameter  $\sigma$  but the mean free path  $\ell$  that limits the validity of the hydrodynamic equations. Therefore, for  $\sigma^{-1} < q < \ell^{-1}$ ,  $S(q, \omega)$  can still be well approximated by three Lorentzians. In this case however,  $A_j$  and  $z_j$  are not determined by hydrodynamic quantities any more.

In summary, for all cases  $q < \ell^{-1}$

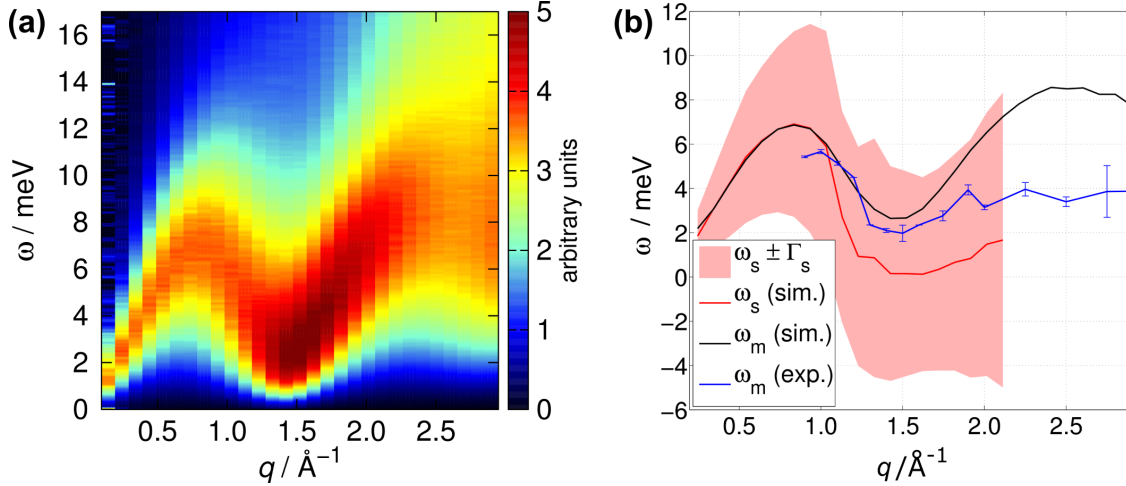
$$\begin{aligned} S(q, \omega) &= \frac{S(q)}{\pi} \sum_{j=-1}^1 \operatorname{Re} \left( \frac{A_j(q)}{i\omega + z_j(q)} \right) \\ &= \frac{S(q)}{\pi} \left\{ A_0 \frac{\Gamma_h}{\omega^2 + \Gamma_h^2} + A_s \left[ \frac{\Gamma_s + (\omega + \omega_s) \tan \phi}{(\omega + \omega_s)^2 + \Gamma_s^2} + \frac{\Gamma_s - (\omega - \omega_s) \tan \phi}{(\omega - \omega_s)^2 + \Gamma_s^2} \right] \right\}, \end{aligned} \quad (7.14)$$

where  $A_0(q)$  is the area of the Rayleigh line and  $A_s(q) = \operatorname{Re} A_{\pm 1}(q)$  the area of the Brillouin lines.  $\Gamma_h(q) = z_0(q)$  and  $\Gamma_s(q) = \operatorname{Re} z_{\pm 1}$  are their corresponding line widths,  $\omega_s = \operatorname{Im} z_1$  is the sound mode frequency, and  $\tan \phi = \operatorname{Im} A_1 / \operatorname{Re} A_1$  determines the asymmetry of the Brillouin lines.

### 7.7.2. Comparison to experiment

The dynamic structure factor  $S(q, \omega)$  of lipid bilayers is usually interpreted in terms of the three effective eigenmode model, i.e. as a sum of a heat and two sound modes (see methods for more details). Unfortunately, the sound modes are often difficult to observe in  $S(q, \omega)$  as they show up only as a soft shoulder in the tails of the heat mode, rather than as a clear excitation maximum. In contrast, in the longitudinal spectrum defined by  $C_\ell(q, \omega) = (\omega^2/q^2)S(q, \omega)$  the multiplication of  $\omega^2$  makes the excitation (sound modes) more easily visible for analysis. Therefore we focus here on  $C_\ell(q, \omega)$  as in the literature on simple fluids (250).

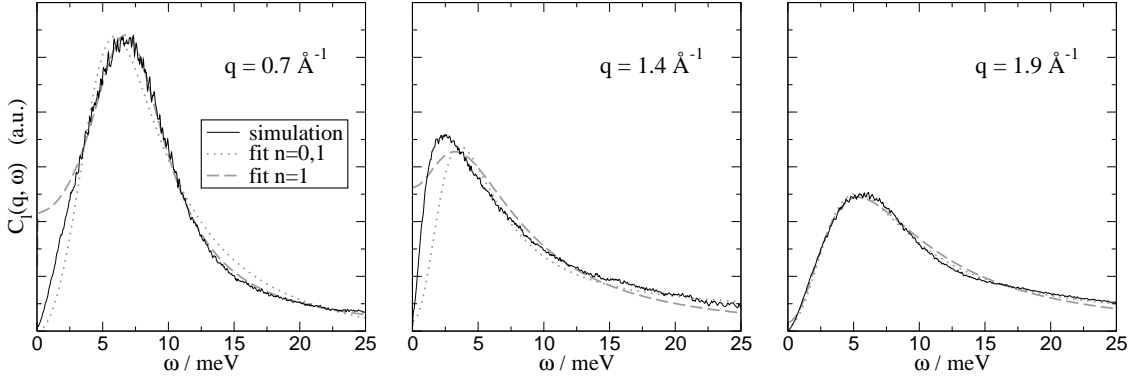
The longitudinal spectrum of the chains of a DMPC bilayer in the fluid  $L_\alpha$  phase extracted from an MD simulation is shown color-coded in fig. 7.8a. In this plot,  $C_\ell(q, \omega)$  has been normalized with respect to its maximum at  $\omega_m$  in the case of a free gas,  $C_{\ell,f}(q, \omega_m) = (2/\pi)^{1/2} e^{-1} v_0/q$ , where  $v_0^2 = k_B T/M$ . As the spectrum is plotted



**Figure 7.8.:** (a) Longitudinal spectrum  $C_\ell(q, \omega)$  normalized with respect to its maximum in the case of a free gas,  $C_{\ell,f}(q, \omega_m)$ . (b) Dispersion curve  $\omega_m(q)$  determined by inelastic neutron scattering (blue) and extracted from the simulation (black), and sound mode frequency  $\omega_s(q)$  determined by a fit to the three effective eigenmode model (red). The width of the sound modes are indicated as a red shaded area of width  $2\Gamma_s(q)$  around  $\omega_s(q)$ .

in arbitrary units the normalization only corresponds to a factor of  $q$ . The mode is dispersive, with a minimum at  $q \sim 1.4 \text{\AA}^{-1}$ , i.e., near the interchain correlation peak of the static SF, and its shape is in agreement with experimental (231, 233) and previous molecular dynamics studies (240).

The maxima of  $C_\ell(q, \omega)$  yield the dispersion curve  $\omega_m(q)$  (fig. 7.8b, black line) which must be distinguished from the sound mode frequency  $\omega_s(q)$  in eq. 7.14. In the same plot, the corresponding dispersion curve determined by neutron scattering (233) is shown as a blue curve, and we find good agreement in the  $q < 2 \text{\AA}^{-1}$  range. For wavelengths shorter than typical interchain distances ( $q \gtrsim 2 \text{\AA}^{-1}$ ), we find higher frequencies in the simulation compared to experiment. Since the force field applied in the simulation has *not* been parameterized to reproduce short wavelength dynamics, the agreement for  $q < 2 \text{\AA}^{-1}$  is an important cross validation for the accuracy of MD simulations. The discrepancy between simulation and experiment in the very short wavelength regime is presumably due to the united atom force field applied for the aliphatic carbons in the simulation. The short range energy landscape of the force field might lack roughness due to the absence of explicit hydrogen or deuterium atoms, rendering the dynamics too rapid.



**Figure 7.9.:** Longitudinal spectrum  $C_\ell(q, \omega)$  (black solid line) for three representative wavenumbers: close to the hydrodynamic limit (top), at the dispersion minimum, i.e. at the reciprocal interchain distance (center), and in the very short wavelength regime (bottom). The corresponding fits to the three effective eigenmode model applying constraint in eq. 7.13 with  $n = 0, 1$  (shaded dashed line), or  $n = 1$  only (shaded dotted line).

### 7.7.3. Does the three-effective-eigenmode model hold?

In order to interpret the spectrum in terms of the three-effective-eigenmode model we have fitted the spectrum extracted from the simulation to eq. 7.14. The fits turned out to be non-trivial. To perform the fit, it is necessary to choose which of the sum rules (eq. 7.13) should be applied as constraints during the fit. The choice is somewhat arbitrary and has also been investigated in experimental studies on short-wavelength modes in liquid argon (248) and neon (249). The spectrum and the corresponding fits are shown for three representative wavenumbers  $q$  in fig. 7.9 as black and shaded lines. We found that the shape and position of the fitted peak agrees best with the simulation results if we apply eq. 7.13 with  $n = 0$  and  $n = 1$ . In this case, we observed some difference between fit and simulation only for small  $\omega$ , indicating that the narrow heat mode decays rather like  $1/\omega$  than  $1/\omega^2$  in the simulation. When additionally applying the sum rule  $n = 2$ , the three eigenmode model is not suitable to fit the data. Applying  $n = 1$  (fig. 7.9, dotted line) or  $n = 0$  only renders an unstable fit which is difficult to reproduce, or the shape of the fitted peak does not match the simulation.

The sound mode frequency  $\omega_s(q)$  resulting from the fit of the three eigenmode model to  $C_\ell(q, \omega)$  from simulation is plotted as a red line in fig. 7.8b. The corresponding line width is given as a red shaded area of width  $2\Gamma_s(q)$  around the  $\omega_s(q)$ . Three regimes should be distinguished:

(i) For  $q \lesssim 1.2 \text{ \AA}^{-1}$  the sound mode frequency  $\omega_s(q)$  equals the dispersion maximum



$\omega_m(q)$  of  $C_\ell(q, \omega)$  indicating that the peak is due to propagation of sound. In the  $q < 0.7 \text{ \AA}^{-1}$  regime,  $\omega_s$  increases approximately linearly with  $q$  as expected in the hydrodynamic limit.

(ii) In the  $q \approx 1.4 \text{ \AA}^{-1}$  regime, i.e. for wavelengths similar to a typical interchain distance, the fit tends to  $\omega_s = 0$ , i.e. to three Lorentzians centered around  $\omega = 0$ . This indicates a dispersion gap and has also been found experimentally in liquid neon (249) and liquid argon (248), and has been proposed theoretically for a system of hard spheres (247). From our results, the propagation of sound is likely to show a dispersion gap also in a lipid bilayer. (It should be noted that the fit does not tend to a propagation gap when applying constraint  $n = 1$  only. In this case, however, the fit was more unstable and its result was unsatisfactory at the dispersion minimum  $q \sim 1.4 \text{ \AA}^{-1}$ , compare fig. 7.9, center.)

(iii) In the range  $q \gtrsim 2.0 \text{ \AA}^{-1}$  the fit turned out to be unstable, and dependent on the fitted  $\omega$  range and the starting points. Therefore, we refrained from analyzing the spectrum with respect to the three eigenmode model for the high  $q$  range.

## 7.8. Conclusion

We have probed the short range order and short wavelength dynamics of a hydrated DMPC bilayer by elastic x-ray diffraction, inelastic neutron scattering, and MD simulations. The combination of experiment and simulation offers a powerful set of tools to investigate the lipid structure and dynamics in both reciprocal and real space. Whereas experiments are essential for force field validation and developments, simulations help to interpret and complement scattering experiments and can, in turn, initiate further experimental studies.

We stress that the methodology developed here in refining the analysis of scattering experiments by MD simulation can be used in a much more general sense, in particular also for multi-component membranes.

We have found that the electron density as well as the SF in lateral direction as determined by simulations agree favorably with experimental results. In particular, position and correlation length of the interchain correlation peak at  $\simeq 1.39 \text{ \AA}$  were well reproduced in all simulations. Good agreement was also found for the reciprocal space mapping of the bilayer representing correlations between tilted lipid tails and their tilt angle distribution.

The head groups were found to display correlation, producing a peak at  $0.72 \text{ \AA}$  in the SF which is, however, hard to observe in x-ray experiments. In order to elucidate structural determinants of the head groups of the bilayer we analyzed RDFs of different components of head groups and identified frequent arrangements of adjacent lipid molecules. We found that although the head groups, in a first approximation, are arranged like a densely packed 2D liquid, their asymmetry induces more complex structural arrangements. These act on the sub-headgroup level and lead to a complex effective energy landscape in which the lipids move. Experimentally, selective deuteration of the molecular components of the head groups and corresponding contrast variation in neutron scattering of oriented membrane stacks may be used to further study this issue.

Simulations of DMPC at different areas per lipid could validate analytical expressions that relate the RDF of the lipid tails to their SF. The most likely interchain distance  $a$ , the mean interchain distance  $a_m$ , and period  $\lambda^{\text{RDF}}$  of the lipid tail RDF can be determined from the SF to very high accuracy. These three quantities proved to be robust properties of the bilayer, and surprisingly, are independent of the area per lipid. In contrast, the interchain correlation length  $\xi$  is linearly dependent on the area per lipid, and might therefore turn out to be a new observable to quantify the area per lipid.

We found the dispersive mode of the lateral spectrum as determined by MD simulations to compare well with inelastic neutron scattering experiments, in particular around the dispersion minimum. At very short wavelengths, the dynamics in the simulations are presumably slightly too rapid due to the united atom force field applied. A quantitative analysis of the spectrum in terms of the three effective eigenmode model is possible, although we do not find agreement over the complete  $q$ - $\omega$  range. This could be taken as an indication that the collective dynamics of the lipid chains cannot be fully described by the theory developed for simple liquids. The issue of a dispersion gap in the propagation of sound remains challenging.

## 8. Summary and conclusions

Tight control of solute permeation across membranes is a common feature of all living organisms. For this purpose evolution has created a tremendous variety of membrane channels including the aquaporin family. Aquaporins facilitate the passive flux of water and other small solutes across biological membranes and have been found in diverse organisms ranging from archae to humans. Since their discovery in 1992 aquaporins have been an increasingly active field of research. A number of high resolution structures could be resolved and permeation experiments elucidated the diverse characteristics and physiological functions of aquaporins.

The present thesis is aimed at a deeper understanding of the function and selectivity of aquaporins, in particular on a molecular and quantitative level. Such insight is crucial to understand the manifold physiological functions of aquaporins and their role in diseases, which is a prerequisite for the development of medical therapies. A detailed insight into solute permeation is particularly important to understand the molecular determinants of channel selectivity and may at some point allow the rational design of functional channels.

**CO<sub>2</sub> permeation through aquaporin-1** The first objective of the present thesis was to reveal the relevance of aquaporin-1 (AQP1) as a gas channel, in particular its putative physiological role as a channel for CO<sub>2</sub>. Using MD simulation, we calculated potentials of mean force (PMFs) for CO<sub>2</sub> permeation across AQP1 embedded in a model membrane of POPE (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine). PMFs provide the probability for the solute to be located at a certain position along the permeation pathway. Therefore they are directly related to the permeability of the channel and help to identify constriction sites along the pathway.

PMFs were derived for every possible pathway of a permeating CO<sub>2</sub> molecule. Since AQPs form tetramers with each monomer contributing one water pore, the possible pathways include the monomeric water pores, a hydrophobic central cavity

along the tetrameric axis as well as the pathways through the lipid bilayer. The simulations revealed a surprisingly large energetic barrier against CO<sub>2</sub> permeation through the AQP1 water pore, rendering AQP1 a rather inefficient CO<sub>2</sub> channel. In contrast, the POPE membrane turned out to be highly permeable to CO<sub>2</sub>. These findings showed that a role of the AQP1 water as a physiological CO<sub>2</sub> channel seems unlikely, at least in membranes similar to POPE.

An intriguing open question in AQP research is the role of the central cavity. For example, it has been speculated that the cavity may function as an ion pore (259, 260). However, it is unclear whether the central cavity is usually open or blocked by an ion or some organic molecule. For an open cavity, the simulations revealed that the permeation of CO<sub>2</sub> through the hydrophobic central cavity is significantly more likely than permeation through one of the water pores. Similar to the water pores however, a significant CO<sub>2</sub> flux through the central cavity is only expected if AQP1 is embedded in a membrane with much lower intrinsic CO<sub>2</sub> permeability than POPE.

The energetic picture of O<sub>2</sub> permeation through AQP1 turned out to be similar to the one of CO<sub>2</sub>. POPE is highly permeable to O<sub>2</sub> as well, whereas O<sub>2</sub> encounters substantial barriers when permeating through AQP1, rendering a physiological role of AQP1 as O<sub>2</sub> channel unlikely.

**The selectivity mechanism of aquaporins and aquaglyceroporins** Having established that both CO<sub>2</sub> and O<sub>2</sub> encounter substantial barriers at the ar/R site of AQP1, the next efforts were undertaken towards a more general picture of the selectivity of aquaporins and aquaglyceroporins. To this end, we derived the energetics of a number of solutes permeating through one typical member of the two AQP subfamily. AQP1 was chosen as a typical AQP water channel and the *E. coli* glycerol facilitator GlpF as a representative for the subfamily of aquaglyceroporins. The selected solutes include urea, glycerol, ammonia, water, CO<sub>2</sub> and O<sub>2</sub>, which differ significantly in hydrophobicity and size.

The main barrier against any solute permeating through AQP1 or GlpF turned out to be located at the narrowest part of the channel, termed aromatic/arginine (ar/R) constriction region. These findings demonstrate the role of the ar/R site as the selectivity filter for uncharged solutes. For the small solutes ammonia, water, CO<sub>2</sub> and O<sub>2</sub> permeating through AQP1, the simulations revealed a remarkable





anti-correlation between solute permeability and hydrophobicity. Surprisingly, not solute–protein but water–protein interactions could be identified as the the molecular mechanism underlying the selectivity filter. Water molecules inside the pore form strong hydrogen bonds mainly to a conserved arginine, but also to a histidine and a cysteine which are located at the ar/R site. We showed that these favorable interactions are reduced upon passage of a solute. The more polar the solute, the stronger it can interact with the polar protein residues, rendering a lower energetic cost to replace a water molecule. The larger solutes urea and glycerol were excluded by size in AQP1. Taken together, the ar/R site of AQP1 forms a filter which allows the passage of small polar solutes.

The ar/R site of GlpF is slightly larger and more hydrophobic than the ar/R site of AQP1. This difference has drastic effects on the selectivity characteristics of the channel as demonstrated by the simulations. The solute hydrophobicity filter present in AQP1 does not apply in GlpF, rendering GlpF generally less selective with lower permeation barriers as compared to AQP1. GlpF is an efficient channel for all considered solutes except for urea. This finding has experimental implications. To address the permeation of small hydrophobic solutes such as CO<sub>2</sub> or NO through AQP1, we suggest to complement permeation experiments on AQP1 with experiments using aquaglyceroporins such as GlpF. If permeation occurs along the AQP1 water channels and not across the lipid bilayer the different ar/R region of GlpF should drastically increase the permeability as compared to AQP1.

**Water flux regulation by a membrane potential** Gating and regulation of membrane channels is a fascinating and active field of research. Ion channels are known to be gated by voltage or ligands (31, 89), whereas several AQPs were reported to be gated or regulated by phosphorylation and pH (188). The present thesis reported for the first time on the voltage-regulation of an AQP channel. In the simulation the permeability of AQP1 could be regulated by ~60% by inverting the membrane potential. As a regulatory mechanism two distinct conformations of the conserved Arg195 in the ar/R site of the channel could be identified, an *up* and a *down* state of Arg195. The membrane potential was found to shift the relative populations of the up and down state towards a more closed channel, rendering a reduced water flux. The understanding of such regulatory mechanisms on a molecular level is a prerequisite to design channels with defined characteristics. Whether point muta-

tions allow to destabilize one of the two states and tune the water permeability of AQP1 remains an interesting topic for future studies.

**AQP1 inhibition by tetraethylammonium** The methodology which was developed for the present thesis to investigate solute permeabilities may be used in a more general sense, for example to identify binding sites and binding free energies of putative inhibitors. As an application with medical implications we investigated the binding of tetraethylammonium (TEA) to AQP1 and its inhibitory effect. Through a combination of different computational approaches the binding site could be identified as a clear energetic minimum. The simulations revealed that TEA does not bind in terms of the “key in lock” model (219), but rather by an induced fit motion of the protein (220) to a TEA molecule which is present in the binding site. In addition, the analysis of the water flux confirmed that TEA is a moderate inhibitor for AQP1. Such combined computational approaches are in general an effective tool to predict and to characterize binding sites for ligand molecules, in particular if co-crystallization is not a routine procedure for the protein of interest.

**Structure and dynamics of lipid membranes** MD simulations apply a number of approximations and must extensively and continuously be validated by comparison to experimental data. Therefore, the present thesis does not only rely on force field models but also aimed to contribute to the ongoing force field development by comparing simulation and experiments in a novel way. A number of structural and short-term dynamic properties were derived from lipid bilayer simulations and compared to results from elastic x-ray and inelastic neutron scattering experiments. Surprisingly good agreement was found, although the lipid force field has not been parameterized with respect to most of the quantities which were compared. Such comparisons are an important cross validation of the general methodology of MD simulations.

In addition to the comparison between simulation and experiments the chapter demonstrated in several ways how simulations can aid the interpretation of experiments, generating a fruitful interplay between simulation and experiments. Moreover, a number of new insights into the structure and dynamics of lipid bilayers were presented such as a simple quantitative relation between the area per lipid and the correlation length of the lipid tail ordering.

# Acknowledgments

First of all, I am very grateful to Dr. Bert de Groot for accepting me into his group at the MPI for Biophysical Chemistry in Göttingen and for encouraging me to study aquaporins. He answered all my numerous questions at the beginning of my PhD project with unlimited patience. He was open to extensive discussions at any time and supported me with valuable and motivating ideas regarding my project, but also gave me the freedom to follow my own ideas. He initiated fruitful collaborations with Prof. Eric Beitz at the University of Kiel and Prof. Tim Salditt at the University of Göttingen. It must be mentioned that he never gave me the impression that he is too busy now to deal with my problems, not a single time in three years!

I also thank Prof. Helmut Grumbüller for his willingness to take over the position of the primary referee for my dissertation, and Prof. Tim Salditt for taking over the position of the second referee.

I thank very much all former and present members of the Department for Theoretical and Computational Biophysics for the friendly welcome into the group, the help with Gromacs, scripting, programming, the interesting scientific discussion. I am particularly grateful to Ira Tremmel for the help in the preparation of the publications and Katja Riebeseel for carefully reading this thesis. Special thanks go to Gerrit Groenhof for his support in the calculation of partial charges, and to Rainer Böckmann for support in the calculation of structure factors.

I thank all members of the MalariaPorin EU consortium for the interesting collaboration and discussions during our meetings in Göteborg, Manchester, and Göttingen. Special thanks go to Prof. Eric Beitz for coordinating the project. I am grateful to Prof. Tim Salditt and Prof. Maikel Rheinstädter for the interesting and fruitful collaboration in the field of lipid bilayers, for providing me with experimental data, and for their extensive help with the preparation of our manuscript. I also thank the Gromacs development team for providing us with the fastest MD code

on earth.

Support by the European Union via grant no. LSHP-CT-2004-012189 and by the Max-Planck-Society is gratefully acknowledged.

A particularly thank you is addressed to Evelin Heinemann, Ansgar Eszermann, Ingo Hoffmann, Martin Fechner, and Oliver Slawik for the administrative and technical support.

This thesis would not exist without ongoing support from my family, my friends, and my girlfriend. Thank You!

# A. Appendix

All partial charges were obtained from electrostatic fitting using the CHELPG procedure to wavefunctions obtained at the RHF/6-31G\* level.

## O<sub>2</sub> topology

- ffoplsnb.itp entry

```
MOLOXY  BOND_MOLOXY 15.99940 -0.11136 A 0.30291 0.5024
OXY_DUM   xxx      0.0      0.22272 D 0.0      0.0
```

- topology o2.itp

```
[ moleculetype ]
; Name nrexcl
O2      3
[ atoms ]
; nr  type  resnr  residu  atom  cgnr  charge  mass
  1  MOLOXY  1      O2      001   1     -0.11136 15.99940
  2  MOLOXY  1      O2      002   1     -0.11136 15.99940
  3  OXY_DUM 1      O2      OD1   1      0.22272  0.0
[ bonds ]
; ai  aj  funct  c0      c1
  1   2   1      0.123   251208.0
[ dummies2 ]
;Dummy  from  funct  a
  3      1   2      1      0.5
[ exclusions ]
1       2       3
2       1       3
```

## CO<sub>2</sub> topology

```
[ moleculetype ]
; Name nrexcl
CO2     3
[ atoms ]
; nr  type  resnr  residue  atom  cgnr  charge  mass
  1  opls_235  1      CO2      CO    1     0.937779  0.
  2  opls_236  1      CO2      OO    1    -0.468889  0.
  3  opls_236  1      CO2      OO    1    -0.468889  0.
  4  opls_112  1      CO2      RM    1      0.0      22.0325
  5  opls_112  1      CO2      RM    1      0.0      22.0325
[ dummies2 ]
1   4   5   1   0.5
2   4   5   1  -0.0862
3   4   5   1  1.0862
```

```
[ constraints ]
4   5   1   0.196
[ exclusions ]
1     2     3     4     5
2     1     3     4     5
3     1     2     4     5
```

## Tetraethylammonium TEA

ffoplsaa.rtp entry for TEA. Topology may be generated with pdb2gmx.

```
; Tetra Ethyl Ammonium (TEA)
[ TEA ]
[ atoms ]
N      opl_940  0.164342  1
CA1    opl_943  0.0491847  2
HA11   opl_140  0.0644464  2
HA12   opl_140  0.0644464  2
CA2    opl_135  -0.266175  3
HA21   opl_140  0.0990041  3
HA22   opl_140  0.0990041  3
HA23   opl_140  0.0990041  3
CB1    opl_943  0.0491847  4
HB11   opl_140  0.0644464  4
HB12   opl_140  0.0644464  4
CB2    opl_135  -0.266175  5
HB21   opl_140  0.0990041  5
HB22   opl_140  0.0990041  5
HB23   opl_140  0.0990041  5
CC1    opl_943  0.0491847  6
HC11   opl_140  0.0644464  6
HC12   opl_140  0.0644464  6
CC2    opl_135  -0.266175  7
HC21   opl_140  0.0990041  7
HC22   opl_140  0.0990041  7
HC23   opl_140  0.0990041  7
CD1    opl_943  0.0491847  8
HD11   opl_140  0.0644464  8
HD12   opl_140  0.0644464  8
CD2    opl_135  -0.266175  9
HD21   opl_140  0.0990041  9
HD22   opl_140  0.0990041  9
HD23   opl_140  0.0990041  9

[ bonds ]
N      CA1
CA1    HA11
CA1    HA12
CA1    CA2
CA2    HA21
CA2    HA22
CA2    HA23
N      CB1
CB1    HB11
CB1    HB12
CB1    CB2
CB2    HB21
CB2    HB22
CB2    HB23
N      CC1
CC1    HC11
CC1    HC12
CC1    CC2
CC2    HC21
CC2    HC22
CC2    HC23
N      CD1
CD1    HD11
CD1    HD12
CD1    CD2
CD2    HD21
CD2    HD22
CD2    HD23

[ dihedrals ]
CA2  CA1  N  CC1
CB2  CB1  N  CD1
CC2  CC1  N  CB1
CD2  CD1  N  CA1
```

# Bibliography

- [1] H. H. Ussing. Transport of electrolytes and water across epithelia. *Harvey Lect*, 59:1–30, 1965.
- [2] R. I. Macey. Transport of water and urea in red blood cells. *Am J Physiol*, 246(3 Pt 1):C195–C203, Mar 1984.
- [3] G. T. Rich, I. Sha'afi, A. Romualdez, and A. K. Solomon. Effect of osmolality on the hydraulic permeability coefficient of red cells. *J Gen Physiol*, 52(6):941–954, Dec 1968.
- [4] A. K. Solomon, B. Chasan, J. A. Dix, M. F. Lukacovic, M. R. Toon, and A. S. Verkman. The aqueous pore in the red cell membrane: band 3 as a channel for anions, cations, nonelectrolytes, and water. *Ann N Y Acad Sci*, 414:97–124, 1983.
- [5] 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.
- [6] L. S. King, D. Kozono, and P. Agre. From structure to disease: the evolving tale of aquaporin biology. *Nature Rev. Molec. Cell Biol.*, 5:687–698, 2004.
- [7] R. Zardoya. Phylogeny and evolution of the major intrinsic protein family. *Biol Cell*, 97(6):397–414, Jun 2005.
- [8] M. Yasui, A. Hazama, T. H. Kwon, S. Nielsen, W. B. Guggino, and P. Agre. Rapid gating and anion permeability of an intracellular aquaporin. *Nature*, 402:184–187, 1999.
- [9] P. Francis, V. Berry, S. Bhattacharya, and A. Moore. Congenital progressive polymorphic cataract caused by a mutation in the major intrinsic protein of the lens, MIP (AQP0). *Br J Ophthalmol*, 84(12):1376–1379, Dec 2000.
- [10] A. Shiels, D. Mackay, S. Bassnett, K. Al-Ghoul, and J. Kuszak. Disruption of lens fiber cell architecture in mice expressing a chimeric aqp0-ltr protein. *FASEB J*, 14(14):2207–2212, Nov 2000.
- [11] D. A. Gorelick, J. Praetorius, T. Tsunenari, S. Nielsen, and P. Agre. Aquaporin-11: a channel protein lacking apparent transport function expressed in brain. *BMC Biochem*, 7:14, 2006.

- [12] T. Itoh, T. Rai, M. Kuwahara, S. B. H. Ko, S. Uchida, S. Sasaki, and K. Ishibashi. Identification of a novel aquaporin, AQP12, expressed in pancreatic acinar cells. *Biochem Biophys Res Commun*, 330(3):832–838, May 2005.
- [13] D. Fu, A. Libson, L. J. Miercke, C. Weitzman, P. Nollert, J. Krucinski, and R. M. Stroud. Structure of a glycerol-conducting channel and the basis for its selectivity. *Science*, 290:481–486, 2000.
- [14] J. K. Lee, D. Kozono, J. Remis, Y. Kitagawa, P. Agre, and R. 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.
- [15] H. Sui, B.-G. Han, J. K. Lee, P. Walian, and B. K. Jap. Structural basis of water-specific transport through the AQP1 water channel. *Nature*, 414:872–878, 2001.
- [16] S. Törnroth-Horsefield, Y. Wang, K. Hedfalk, U. Johanson, M. Karlsson, E. Tajkhorshid, R. Neutze, and P. Kjellbom. Structural mechanism of plant aquaporin gating. *Nature*, 439(7077):688–694, Feb 2006.
- [17] W. E. C. Harries, D. Akhavan, L. J. W. Miercke, S. Khademi, and R. M. Stroud. The channel architecture of aquaporin 0 at a 2.2-Å resolution. *Proc. Natl. Acad. Sci. USA*, 101:14045–14050, 2004.
- [18] Y. Hiroaki, K. Tani, A. Kamegawa, N. Gyobu, K. Nishikawa, H. Suzuki, T. Walz, S. Sasaki, K. Mitsuoka, K. Kimura, A. Mizoguchi, and Y. Fujiyoshi. Implications of the aquaporin-4 structure on array formation and cell adhesion. *J. Mol. Biol.*, 355:625–639, 2006.
- [19] D. F. Savage, P. F. Egea, Y. Robles-Colmenares, J. D. O’Connell III, and R. M. Stroud. Architecture and selectivity in aquaporins: 2.5 Å x-ray structure of aquaporin Z. *PLOS Biology*, 1:e72, 2003.
- [20] T. Walz, T. Hirai, K. Murata, J. B. Heymann, K. Mitsuoka, Y. Fujiyoshi, B. L. Smith, P. Agre, and A. Engel. The three-dimensional structure of Aquaporin-1. *Nature*, 387:624–627, 1997.
- [21] K. Mitsuoka, K. Murata, T. Walz, T. Hirai, P. Agre, J. B. Heymann, A. Engel, and Y. Fujiyoshi. The structure of Aquaporin-1 at 4.5-Å resolution reveals short  $\alpha$ -helices in the center of the monomer. *J. Struct. Biol.*, 128:34–43, 1999.
- [22] B. L. de Groot, J. B. Heymann, A. Engel, K. Mitsuoka, Y. Fujiyoshi, and H. Grubmüller. The fold of human Aquaporin 1. *J. Mol. Biol.*, 300:987–994, 2000.
- [23] H. Stahlberg, T. Braun, B. de Groot, A. Philippsen, M. J. Borgnia, P. Agre, W. Kühlbrandt, and A. Engel. The 6.9 Å structure of GlpF: A basis for homology modelling of the glycerol channel from *E. coli*. *J. Struct. Biol.*, 132:133–141, 2000.





- [24] P. Nollert, W. E. Harries, D. Fu, L. J. Miercke, and R. M. Stroud. Atomic structure of a glycerol channel and implications for substrate permeation in aqua(glycero)porins. *FEBS Lett*, 504(3):112–117, Aug 2001.
- [25] G. M. Pao, L. F. Wu, K. D. Johnson, H. Höfte, M. J. Chrispeels, G. Sweet, N. N. Sandal, and M. H. Saier. Evolution of the MIP family of integral membrane transport proteins. *Mol Microbiol*, 5(1):33–37, Jan 1991.
- [26] G. J. Wistow, M. M. Pisano, and A. B. Chepelinsky. Tandem sequence repeats in transmembrane channel proteins. *Trends. Biochem. Sci.*, 16:170–171, 1991.
- [27] T. Gonen and T. Walz. The structure of aquaporins. *Q Rev Biophys*, 39(4):361–396, Nov 2006.
- [28] T. Walz, B. L. Smith, M. L. Zeidel, A. Engel, and P. Agre. Biologically active two-dimensional crystals of aquaporin chip. *J. Biol. Chem.*, 269:1583–1586, 1994.
- [29] M. L. Zeidel, S. V. Ambudkar, B. L. Smith, and P. Agre. Reconstitution of functional water channels in liposomes containing purified red-cell CHIP28 protein. *Biochemistry*, 31:7436–7440, 1992.
- [30] B. L. de Groot and H. Grubmüller. Water permeation across biological membranes: Mechanism and dynamics of Aquaporin-1 and GlpF. *Science*, 294:2353–2357, 2001.
- [31] J. M. Berg, J. L. Tymoczko, and L. Stryer. *Biochemistry*. W. H. Freeman and co., New York, 5th edition, 2002.
- [32] D. Voet and J. G. Voet. *Biochemistry*. Wiley, New York, 2004.
- [33] H. Kandori, Y. Yamazaki, J. Sasaki, R. Needleman, J. K. Lanyi, and A. Maeda. Water-mediated proton-transfer in proteins - an FTIR study of bacteriorhodopsin. *J. Am. Chem. Soc.*, 117:2118–2119, 1995.
- [34] S. B. Hladky and D. A. Haydon. Ion transfer across lipid membranes in the presence of gramicidin A. *Biochim. Biophys. Acta*, 274:294–312, 1972.
- [35] M. Akeson and D. W. Deamer. Proton conductance by the gramicidin water wire. *Biophys. J.*, 60:101–109, 1991.
- [36] C. J. T. de Grotthuss. Sur la décomposition de l'eau et des corps qu'elle tient en dissolution à l'aide de l'électricité galvanique. *Ann. Chim.*, LVIII:54–74, 1806.
- [37] C. Dellago, M. M. Naor, and G. Hummer. Proton transport through water-filled carbon nanotubes. *Phys. Rev. Lett.*, 90:105902–1–105902–4, 2003.

- [38] E. Tajkhorshid, P. Nollert, M. Ø. Jensen, L. J. W. Miercke, J. O'Connell, R. M. Stroud, and K. Schulten. Control of the selectivity of the aquaporin water channel family by global orientational tuning. *Science*, 296:525–530, 2002.
- [39] A. Burykin and A. Warshel. What really prevents proton transport through aquaporin? Charge self-energy versus proton wire proposals. *Biophys. J.*, 85:3696–3706, 2003.
- [40] A. Burykin and A. Warshel. On the origin of the electrostatic barrier for proton transport in aquaporin. *FEBS Lett.*, 570:41–46, 2004.
- [41] N. Chakrabarti, B. Roux, and R. Pomes. Structural determinants of proton blockage in aquaporins. *J. Mol. Biol.*, 343:493–510, 2004.
- [42] N. Chakrabarti, E. Tajkhorshid, B. Roux, and R. Pomes. Molecular basis of proton blockage in aquaporins. *Structure*, 12:65–74, 2004.
- [43] B. L. de Groot, T. Frigato, V. Helms, and H. Grubmüller. The mechanism of proton exclusion in the aquaporin-1 water channel. *J. Mol. Biol.*, 333:279–293, 2003.
- [44] B. Eisenberg. Why can't protons move through water channels? *Biophys. J.*, 85:3427–3428, 2003.
- [45] B. L. de Groot and H. Grubmüller. The dynamics and energetics of water permeation and proton exclusion in aquaporins. *Curr. Opin. Struct. Biol.*, 15:176–183, 2005.
- [46] M. Kato, A. V. Pisliakov, and A. Warshel. The barrier for proton transport in aquaporins as a challenge for electrostatic models: the role of protein relaxation in mutational calculations. *Proteins*, 64(4):829–844, Sep 2006.
- [47] D. Zhang, L. Vetrivel, and A. S. Verkman. Aquaporin deletion in mice reduces intraocular pressure and aqueous fluid production. *J Gen Physiol*, 119(6):561–569, Jun 2002.
- [48] T. H. Kwon, H. Hager, L. N. Nejsum, M. L. Andersen, J. Fr̃̃kiaer, and S. Nielsen. Physiology and pathophysiology of renal aquaporins. *Semin Nephrol*, 21(3):231–238, May 2001.
- [49] M. Endo, R. K. Jain, B. Witwer, and D. Brown. Water channel (aquaporin 1) expression and distribution in mammary carcinomas and glioblastomas. *Microvasc Res*, 58(2):89–98, Sep 1999.
- [50] X. C. Feng, H. W. Gao, C. Y. He, and T. Ma. Defective tumor angiogenesis and retarded tumor growth in aquaporin-1 knockout mice. *Prog Biochem Biophys*, 32:310–313, 2005.
- [51] S. Saadoun, M. C. Papadopoulos, M. Hara-Chikuma, and A. S. Verkman. Impairment of angiogenesis and cell migration by targeted aquaporin-1 gene disruption. *Nature*, 434(7034):786–792, Apr 2005.



- [52] M. C. Papadopoulos, S. Saadoun, D. K. Binder, G. T. Manley, S. Krishna, and A. S. Verkman. Molecular mechanisms of brain tumor edema. *Neuroscience*, 129(4):1011–1020, 2004.
- [53] G. T. Manley, M. Fujimura, T. Ma, N. Noshita, F. Filiz, A. W. Bollen, P. Chan, and A. S. Verkman. Aquaporin-4 deletion in mice reduces brain edema after acute water intoxication and ischemic stroke. *Nature Med*, 6(2):159–163, Feb 2000.
- [54] M. C. Papadopoulos, G. T. Manley, S. Krishna, and A. S. Verkman. Aquaporin-4 facilitates reabsorption of excess fluid in vasogenic brain edema. *FASEB J*, 18:1291, 2004.
- [55] A. Vacca, A. Frigeri, D. Ribatti, G. P. Nicchia, B. Nico, R. Ria, M. Svelto, and F. Dammacco. Microvessel overexpression of aquaporin 1 parallels bone marrow angiogenesis in patients with active multiple myeloma. *Br J Haematol*, 113(2):415–421, May 2001.
- [56] P. Agre and D. Kozono. Aquaporin water channels: molecular mechanisms for human diseases. *FEBS Lett*, 555(1):72–78, Nov 2003.
- [57] N. A. Castle. Aquaporins as targets for drug discovery. *Drug Discovery Today*, 10:485–493, 2005.
- [58] F. Wang, X. chao Feng, Y. ming Li, H. Yang, and T. hui Ma. Aquaporins as potential drug targets. *Acta Pharmacol. Sinica*, 27:395–401, 2006.
- [59] N. L. Nakhoul, B. A. Davis, M. F. Romero, and W. F. Boron. Effect of expressing the water channel aquaporin-1 on the CO<sub>2</sub> permeability of *Xenopus* oocytes. *Am. J. Physiol. Cell Physiol.*, 274:C543–C548, 1998.
- [60] L. M. Holm, T. P. Jahn, A. L. B. Møller, J. K. Schjoerring, D. Ferri, D. A. Klaerke, and T. Zeuthen. NH<sub>3</sub> and NH<sub>4</sub><sup>+</sup> permeability in aquaporin-expressing xenopus oocytes. *Pflugers Arch*, 450(6):415–428, Sep 2005.
- [61] T. P. Jahn, A. L. B. Møller, T. Zeuthen, L. M. Holm, D. A. Klaerke, B. Mohsin, W. Kühlbrandt, and J. K. Schjoerring. Aquaporin homologues in plants and mammals transport ammonia. *FEBS Lett*, 574(1-3):31–36, Sep 2004.
- [62] M. Hara-Chikuma and A. S. Verkman. Physiological roles of glycerol-transporting aquaporins: the aquaglyceroporins. *Cell Mol Life Sci*, 63(12):1386–1392, Jun 2006.
- [63] I. Johansson, M. Karlsson, V. K. Shukla, M. J. Chrispeels, C. Larsson, and P. Kjellbom. Water transport activity of the plasma membrane aquaporin PM28A is regulated by phosphorylation. *Plant Cell*, 10(3):451–459, Mar 1998.

- [64] M. J. Borgnia and P. Agre. Reconstitution and functional comparison of purified GlpF and AqpZ, the glycerol and water channels from *Escherichia coli*. *Proc. Natl. Acad. Sci. USA*, 98:2888–2893, 2001.
- [65] A. S. Verkman. Does aquaporin-1 pass gas? An opposing view. *J. Physiol.*, 524.1:31, 2002.
- [66] M. Baum and R. Quigley. Proximal tubule water transport-lessons from aquaporin knockout mice. *Am J Physiol Renal Physiol*, 289(6):F1193–F1194, Dec 2005.
- [67] V. Vallon, A. S. Verkman, and J. Schnermann. Luminal hypotonicity in proximal tubules of aquaporin-1-knockout mice. *Am J Physiol Renal Physiol*, 278(6):F1030–F1033, Jun 2000.
- [68] G. J. Cooper and W. F. Boron. Effect of PCMBs on CO<sub>2</sub> permeability of *Xenopus* oocytes expressing aquaporin 1 or its C189S mutant. *Amer. J. Physiol.*, 275:C1481–1486, 1998.
- [69] G. J. Cooper, Y. Zhou, P. Bouyer, I. I. Grichtchenko, and W. F. Boron. Transport of volatile solutes through AQP1. *J. Physiol.*, 524.1:17–29, 2002.
- [70] N. Uehlein, C. Lovisolo, F. Siefritz, and R. Kaldenhoff. The tobacco aquaporin NtAQP1 is a membrane CO<sub>2</sub> pore with physiological functions. *Nature*, 425:734–737, 2003.
- [71] B. Yang, N. Fukuda, A. van Hoek, M. A. Matthay, T. Ma, and A. S. Verkman. Carbon dioxide permeability of aquaporin-1 measured in erythrocytes and lung of aquaporin-1 null mice and in reconstituted proteoliposomes. *J. Biol. Chem.*, 275:2686–2692, 2000.
- [72] B. Wu and E. Beitz. Aquaporins with selectivity for unconventional permeants. *Cell Mol Life Sci*, 64(18):2413–2421, Sep 2007.
- [73] G. P. Bienert, A. L. B. Møller, K. A. Kristiansen, A. Schulz, I. M. Møller, J. K. Schjoerring, and T. P. Jahn. Specific aquaporins facilitate the diffusion of hydrogen peroxide across membranes. *J Biol Chem*, 282(2):1183–1192, Jan 2007.
- [74] M. Herrera, N. J. Hong, and J. L. Garvin. Aquaporin-1 transports no across cell membranes. *Hypertension*, 48(1):157–164, Jul 2006.
- [75] I. I. Ivanov, A. V. Loktyushkin, R. A. Guskova, N. S. Vasilev, G. E. Fedorov, and A. B. Rubin. Oxygen channels of erythrocyte membrane. *Doklady Biochem. Biophys.*, 414:137–140, 2007.
- [76] Z. Liu, J. Shen, J. M. Carbrey, R. Mukhopadhyay, P. Agre, and B. P. Rosen. Arsenite transport by mammalian aquaglyceroporins AQP7 and AQP9. *Proc Natl Acad Sci U S A*, 99(9):6053–6058, Apr 2002.
- [77] J. F. Ma, K. Tamai, N. Yamaji, N. Mitani, S. Konishi, M. Katsuhara, M. Ishiguro, Y. Murata, and M. Yano. A silicon transporter in rice. *Nature*, 440(7084):688–691, Mar 2006.



- [78] O. I. Sanders, C. Rensing, M. Kuroda, B. Mitra, and B. P. Rosen. Antimonite is accumulated by the glycerol facilitator GlpF in *Escherichia coli*. *J Bacteriol*, 179(10):3365–3367, May 1997.
- [79] K. B. Heller, E. C. Lin, and T. H. Wilson. Substrate-specificity and transport-properties of the glycerol facilitator of *Escherichia coli*. *J. Bacteriol.*, 144:274–278, 1980.
- [80] S. Nielsen, C. L. Chou, D. Marples, E. I. Christensen, B. K. Kishore, and M. A. Knepper. Vasopressin increases water permeability of kidney collecting duct by inducing translocation of aquaporin-cd water channels to plasma membrane. *Proc Natl Acad Sci U S A*, 92(4):1013–1017, Feb 1995.
- [81] P. M. Deen and D. Brown. Trafficking of native and mutant mammalian MIP proteins. In S. Hohmann, S. Nielsen, and P. Agre, editors, *Aquaporins*, volume 51, pages 235–276. Academic Press, San Diego, 2001.
- [82] R. A. Marinelli, L. Pham, P. Agre, and N. F. LaRusso. Secretin promotes osmotic water transport in rat cholangiocytes by increasing aquaporin-1 water channels in plasma membrane. evidence for a secretin-induced vesicular translocation of aquaporin-1. *J Biol Chem*, 272(20):12984–12988, May 1997.
- [83] R. A. Marinelli, P. S. Tietz, L. D. Pham, L. Rueckert, P. Agre, and N. F. LaRusso. Secretin induces the apical insertion of aquaporin-1 water channels in rat cholangiocytes. *Am J Physiol*, 276(1 Pt 1):G280–G286, Jan 1999.
- [84] Z. Han and R. V. Patil. Protein kinase a-dependent phosphorylation of aquaporin-1. *Biochem Biophys Res Commun*, 273(1):328–332, Jun 2000.
- [85] T. Zeuthen and D. A. Klaerke. Transport of water and glycerol in aquaporin 3 is gated by H(+). *J Biol Chem*, 274(31):21631–21636, Jul 1999.
- [86] K. L. Németh-Cahalan and J. E. Hall. pH and calcium regulate the water permeability of aquaporin 0. *J Biol Chem*, 275(10):6777–6782, Mar 2000.
- [87] I. Johansson, C. Larsson, B. Ek, and P. Kjellbom. The major integral proteins of spinach leaf plasma membranes are putative aquaporins and are phosphorylated in response to  $\text{Ca}^{2+}$  and apoplastic water potential. *Plant Cell*, 8(7):1181–1191, Jul 1996.
- [88] C. Tournaire-Roux, M. Sutka, H. Javot, E. Gout, P. Gerbeau, D.-T. Luu, R. Bligny, and C. Maurel. Cytosolic pH regulates root water transport during anoxic stress through gating of aquaporins. *Nature*, 425(6956):393–397, Sep 2003.
- [89] F. Bezanilla. Voltage-gated ion channels. *IEEE Trans Nanobioscience*, 4(1):34–48, Mar 2005.

- [90] G. Yellen. The moving parts of voltage-gated ion channels. *Q Rev Biophys*, 31(3):239–295, Aug 1998.
- [91] H. J. C. Berendsen. *Simulating the Physical World: Hierarchical Modeling from Quantum Mechanics to Fluid Dynamics*. Cambridge University Press, 2007.
- [92] M. Born and R. Oppenheimer. Quantentheorie der Molekeln. *Annalen der Physik*, 84:457–484, 1927.
- [93] C. Eckart. Some studies concerning rotating axes and polyatomic molecules. *Physical Review*, 46:383–387, 1935.
- [94] P. Ehrenfest. Bemerkung über die angenäherte Gültigkeit der klassischen Mechanik innerhalb der Quantenmechanik. *Zeitschrift für Physik A*, 45:455–457, 1927.
- [95] E. Joos, H. Zeh, C. Kiefer, D. Giulini, J. Kupsch, and I. Stamatescu. *Decoherence and the Appearance of a Classical World in Quantum Theory*. Springer, Berlin, edition 2. A, 2. a edition, 2007.
- [96] W. L. Jorgensen and J. Tirado-Rives. Potential energy functions for atomic-level simulations of water and organic and biomolecular systems. *Proc Natl Acad Sci U S A*, 102(19):6665–6670, May 2005.
- [97] D. Van der Spoel, E. Lindahl, B. Hess, A. R. Van Buuren, E. Apol, P. J. Meulenhoff, D. P. Tieleman, A. L. T. Sijbers, K. A. Feenstra, R. van Drunen, and H. J. C. Berendsen. *Gromacs User Manual version 3.2*, 2004. [www.gromacs.org](http://www.gromacs.org).
- [98] W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.*, 118:11225–11236, 1996.
- [99] G. A. Kaminski, R. A. Friesner, J. Tirado-Rives, and W. L. Jorgensen. Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J. Phys. Chem. B*, 105:6474–6487, 2001.
- [100] C. Oostenbrink, A. Villa, A. E. Mark, and W. 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, Oct 2004.
- [101] W. Van Gunsteren, S. Billeter, A. Eising, P. Hünenberger, P. Krüger, A. Mark, W. Scott, and I. Tironi. *Biomolecular simulation: the GROMOS96 manual and user guide*. Biomos b.v., Zürich, Groningen, 1996.
- [102] J. W. Ponder and D. A. Case. Force fields for protein simulations. *Adv Protein Chem*, 66:27–85, 2003.



- [103] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. CHARMM: a program for macromolecular energy, minimization, and dynamics calculation. *J. Comp. Chem.*, 4:187–217, 1983.
- [104] A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchmir, 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. Wiorcikiewicz-Kuczera, D. Yin, and M. Karplus. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B*, 102:3586–3616, 1998.
- [105] H. J. C. Berendsen, D. van der Spoel, and R. van Drunen. GROMACS: A message-passing parallel molecular dynamics implementation. *Comp. Phys. Commun.*, 91:43–56, 1995.
- [106] D. Van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen. GROMACS: Fast, flexible and free. *J. Comp. Chem.*, 26:701–1719, 2005.
- [107] R. Hockney and S. Goel. Quiet high-resolution computer models of a plasma. *J. Comput. Phys.*, 14:148–158, 1974.
- [108] L. Verlet. Computer “experiments” on classical fluids. I. thermodynamical properties of Lennard-Jones molecules. *Phys. Rev.* 159, 98 - 103 (1967), 159:98–103, 1967.
- [109] B. Hess, H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije. LINCS: A linear constraint solver for molecular simulations. *J. Comp. Chem.*, 18:1463–1472, 1997.
- [110] S. Miyamoto and P. A. Kollman. SETTLE: An analytical version of the SHAKE and RATTLE algorithms for rigid water models. *J. Comp. Chem.*, 13:952–962, 1992.
- [111] M. Saito. Molecular dynamics simulations of proteins in solution: Artifacts caused by the cutoff approximation. *J. Chem. Phys.*, 101:4055–4061, 1994.
- [112] M. Saito. Molecular dynamic/free energy study of a protein in solution with all degrees of freedom and long-range coulomb interactions. *J. Phys. Chem.*, 99:17043–17048, 1995.
- [113] P. P. Ewald. Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Ann. Phys.*, 64:253–287, 1921.
- [114] T. Darden, D. York, and L. Pedersen. Particle mesh Ewald: an  $N \cdot \log(N)$  method for Ewald sums in large systems. *J. Chem. Phys.*, 98:10089–10092, 1993.
- [115] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen. A smooth particle mesh ewald potential. *J. Chem. Phys.*, 103:8577–8592, 1995.
- [116] H. J. C. Berendsen, J. P. M. Postma, A. DiNola, and J. R. Haak. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, 81:3684–3690, 1984.

- [117] T. Morishita. Fluctuation formulas in molecular-dynamics simulations with the weak coupling heat bath. *J. Chem. Phys.*, 113:2976, 2000.
- [118] S. Nosé and M. L. Klein. Constant-temperature-constant-pressure molecular-dynamics calculations for molecular solids: Application to solid nitrogen at high pressure. *Phys Rev B Condens Matter*, 33(1):339–342, Jan 1986.
- [119] M. Parrinello and A. Rahman. Polymorphic transitions in single crystals: A new molecular dynamics method. *J Appl Phys*, 52:7182–7190, 1981.
- [120] W. G. Hoover. Canonical dynamics: Equilibrium phase-space distributions. *Phys Rev A*, 31(3):1695–1697, Mar 1985.
- [121] S. Nosé. A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.*, 52:255–268, 1984.
- [122] J. G. Kirkwood. Statistical mechanics of fluid mixtures. *J. Chem. Phys.*, 3:300–313, 1935.
- [123] W. Greiner and H. Stocker. *Thermodynamics and Statistical Mechanics*. Springer, New York, 1997.
- [124] P. M. Chaikin and T. C. Lubensky. *Principles of condensed matter physics*. Cambridge University Press, Cambridge, 2000.
- [125] I. Kosztin, B. Barz, and L. Janosi. Calculating potentials of mean force and diffusion coefficients from nonequilibrium processes without Jarzynski’s equality. *J Chem Phys*, 124(6):64106, Feb 2006.
- [126] G. M. Torrie and J. P. Valleau. Monte Carlo free energy estimates using non-Boltzmann sampling: Application to the sub-critical Lennard-Jones fluid. *Chemical Physics Letters*, 28:578–581, October 1974.
- [127] J. P. Valleau and G. M. Torrie. A guide for Monte Carlo for statistical mechanics. In *Statistical Mechanics, part A*, pages 169–194. Plenum Press, New York, 1977.
- [128] H. Grubmüller, B. Heymann, and P. Tavan. Ligand binding: Molecular mechanics calculation of the streptavidin-biotin rupture force. *Science*, 271:997–999, 1996.
- [129] B. Isralewitz, J. Baudry, J. Gullingsrud, D. Kosztin, and K. Schulten. Steered molecular dynamics investigations of protein function. *J Mol Graph Model*, 19(1):13–25, 2001.
- [130] G. Hummer and A. Szabo. Free energy reconstruction from non-equilibrium single-molecule pulling experiments. *Proc. Natl. Acad. Sci. USA*, 98:3658–3661, 2001.
- [131] C. Jarzynski. Equilibrium free-energy differences from nonequilibrium measurements: A master-equation approach. *Phys. Rev. E*, 56:5018–5035, 1997.





- [132] C. Jarzynski. Nonequilibrium equality for free energy difference. *Phys. Rev. Lett.*, 78:2690–2693, 1997.
- [133] G. Hummer. Fast-growth thermodynamic integration: Error and efficiency analysis. *J Chem Phys*, 114:7330–7337, 2001.
- [134] M. Ø. Jensen, S. Park, E. Tajkhorshid, and K. Schulten. Energetics of glycerol conduction through aquaglyceroporin GlpF. *Proc. Natl. Acad. Sci. USA*, 99:6731–6736, 2002.
- [135] D. J. Evans and D. J. Searles. Equilibrium microstates which generate second law violating steady states. *Phys. Rev. E*, 50(2):1645–1648, Aug 1994.
- [136] D. J. Evans and D. J. Searles. Causality, response theory, and the second law of thermodynamics. *Phys. Rev. E*, 53(6):5808–5815, Jun 1996.
- [137] G. E. Crooks. Path-ensemble averages in systems driven far from equilibrium. *Phys. Rev. E*, 61(3):2361–2366, Mar 2000.
- [138] I. McDonalda and J. Rasaiaha. Monte carlo simulation of the average force between two ions in a stockmayer solvent. *Chem Phys Lett*, 34:382–386, 1975.
- [139] R. W. Zwanzig. High-temperature equation of state by a perturbation method. 1. Nonpolar gases. *J. Chem. Phys.*, 22:1420–1426, 1954.
- [140] J. Valleau and S. Whittington. A guide to monte carlo for statistical mechanics: Highways in modern theoretical chemistry. In B. Berne, editor, *Modern Theoretical Chemistry*. Plenum, New York, 1977.
- [141] T. P. Straatsma, H. J. C. Berendsen, and J. P. M. Postma. Free energy of hydrophobic hydration: A molecular dynamics study of noble gases in water. *J. Chem. Phys.*, 85:6720–6727, 1986.
- [142] T. Straatsma, , and J. A. McCammon. Computational alchemy. *Ann. Rev. Phys. Chem.*, 43:407–435, 1992.
- [143] B. Widom. Some topics in theory of fluids. *J Chem Phys*, 39:2808, 1963.
- [144] B. Roux. The calculation of the potential of mean force using computer simulations. *Comp. Phys. Comm.*, 91:275–282, 1995.
- [145] J. Shen and J. McCammon. Molecular dynamics simulation of superoxide interacting with superoxide dismutase. *Chemical physics*, 158:191–198, 1991.
- [146] C. Haydock, J. C. Sharp, and F. G. Prendergast. Tryptophan-47 rotational isomerization in variant-3 scorpion neurotoxin. a combination thermodynamic perturbation and umbrella sampling study. *Biophys J*, 57(6):1269–1279, Jun 1990.

- [147] S. Kumar, D. Bouzida, R. H. Swendsen, P. A. Kollman, and J. M. Rosenberg. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comp. Chem.*, 13:1011–1021, 1992.
- [148] Ferrenberg and Swendsen. Optimized monte carlo data analysis. *Phys Rev Lett*, 63(12):1195–1198, Sep 1989.
- [149] M. H. Abraham, G. J. Buist, P. L. Grellier, R. A. McGill, R. M. Doherty, M. J. Kamlet, R. W. Taft, and S. G. Maroldo. Solubility properties in polymers and biological media. ii. a new method for the characterisation of the adsorption of gases and vapours on solids. *J Chromatogr*, 409:15–27, Nov 1987.
- [150] A. Finkelstein. Water and nonelectrolyte permeability of lipid bilayer membranes. *J. Gen. Physiol.*, 68(2):127–135, Aug 1976.
- [151] S. A. Simon and J. Gutknecht. Solubility of carbon dioxide in lipid bilayer membranes and organic solvents. *Biochim. Biophys. Acta*, 596(3):352–358, Mar 1980.
- [152] A. Walter and J. Gutknecht. Permeability of small nonelectrolytes through lipid bilayer membranes. *J. Membr. Biol.*, 90(3):207–217, 1986.
- [153] D. M. Pollock. Chronic studies on the interaction between nitric oxide and endothelin in cardiovascular and renal function. *Clin. Exp. Pharmacol. Physiol.*, 26(3):258–261, Mar 1999.
- [154] B. C. Kone. Nitric oxide synthesis in the kidney: isoforms, biosynthesis, and functions in health. *Semin Nephrol*, 24(4):299–315, Jul 2004.
- [155] S. H. Snyder and D. S. Bredt. Nitric oxide as a neuronal messenger. *Trends Pharmacol Sci*, 12(4):125–128, Apr 1991.
- [156] E. Overton. Über die osmotischen Eigenschaften der lebenden Pflanzen- und Tierzelle. *Vierteljahresschr. Naturforsch. Ges. Zürich*, 40:159–201, 1895.
- [157] J. L. Garvin, M. B. Burg, and M. A. Knepper. Active  $\text{nh}_4^+$  absorption by the thick ascending limb. *Am J Physiol*, 255(1 Pt 2):F57–F65, Jul 1988.
- [158] S. K. Singh, H. J. Binder, J. P. Geibel, and W. F. Boron. An apical permeability barrier to  $\text{NH}_3/\text{NH}_4^+$  in isolated, perfused colonic crypts. *Proc. Natl. Acad. Sci. USA*, 92(25):11573–11577, Dec 1995.
- [159] N. A. Priver, E. C. Rabon, and M. L. Zeidel. Apical membrane of the gastric parietal cell: water, proton, and nonelectrolyte permeabilities. *Biochemistry*, 32(10):2459–2468, Mar 1993.



- [160] S. J. Waisbren, J. P. Geibel, I. M. Modlin, and W. F. Boron. Unusual permeability properties of gastric gland cells. *Nature*, 368(6469):332–335, Mar 1994.
- [161] X. Fang, B. Yang, M. A. Matthay, and A. S. Verkman. Evidence against aquaporin-1-dependent CO<sub>2</sub> permeability in lung and kidney. *J. Physiol.*, 542:63–69, 2002.
- [162] B. Yang, D. Zhao, E. Solenov, and A. S. Verkman. Evidence from knockout mice against physiologically significant aquaporin 8-facilitated ammonia transport. *Am J Physiol Cell Physiol*, 291(3):C417–C423, Sep 2006.
- [163] M. E. Blank and H. Ehmke. Aquaporin-1 and HCO<sub>3</sub><sup>-</sup>-Cl<sup>-</sup> transporter-mediated transport of CO<sub>2</sub> across the human erythrocyte membrane. *J. Physiol.*, 550.2:419–429, 2003.
- [164] G. V. R. Prasad, L. A. Coury, F. Finn, and M. L. Zeidel. Reconstituted aquaporin 1 water channels transport CO<sub>2</sub> across membranes. *J. Biol. Chem.*, 273:33123–33126, 1998.
- [165] J. Gutknecht, M. A. Bisson, and F. C. Tosteson. Diffusion of carbon dioxide through lipid bilayer membranes: effects of carbonic anhydrase, bicarbonate, and unstirred layers. *J. Gen. Physiol.*, 69:779–794, 1977.
- [166] S. J. Waisbren, J. P. Geibel, I. M. Modlin, and W. F. Boron. Unusual permeability properties of gastric gland cells. *Nature*, 368:332–335, 1994.
- [167] G. Vriend. WHAT IF: a molecular modeling and drug design program. *J. Mol. Graphics*, 8:52–56, 1990.
- [168] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, 79:926–935, 1983.
- [169] 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.*, 72:2002–2013, 1997.
- [170] E. Lindahl, B. Hess, and D. Van der Spoel. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *J. Mol. Model.*, 7:306–317, 2001.
- [171] R. E. Forster, G. Gros, L. Lin, Y. Ono, and M. Wunder. The effect of 4,4'-diisothiocyanatostilbene-2,2'-disulfonate on CO<sub>2</sub> permeability of the red blood cell membrane. *Proc. Natl. Acad. Sci. USA*, 95:15815–15820, 1998.
- [172] M. Ø. Jensen and O. G. Mouritsen. Single-channel water permeabilities of Escherichia coli aquaporins AqpZ and GlpF. *Biophys. J.*, 90:2270–2284, 2006.
- [173] J. Jiang, B. V. Daniels, and D. Fu. Crystal structure of aquaporin z tetramer reveals both open and closed water-conducting channels. *J. Biol. Chem.*, 281:454–460, 2006.

- [174] I. I. Ivanov, G. E. Fedorov, R. A. Gus'kova, K. I. Ivanov, and A. B. Rubin. Permeability of lipid membranes to dioxygen. *Biochem Biophys Res Commun*, 322(3):746–750, Sep 2004.
- [175] A. Ligeza, A. N. Tikhonov, J. S. Hyde, and W. K. Subczynski. Oxygen permeability of thylakoid membranes: electron paramagnetic resonance spin labeling study. *Biochim Biophys Acta*, 1365(3):453–463, Jul 1998.
- [176] J. Widomska, M. Raguz, and W. K. Subczynski. Oxygen permeability of the lipid bilayer membrane made of calf lens lipids. *Biochim Biophys Acta*, Jun 2007.
- [177] W. K. Subczynski, J. S. Hyde, and A. Kusumi. Oxygen permeability of phosphatidylcholine–cholesterol membranes. *Proc. Natl. Acad. Sci. USA*, 86(12):4474–4478, Jun 1989.
- [178] Y. Wang, J. Cohen, W. F. Boron, K. Schulten, and E. Tajkhorshid. Exploring gas permeability of cellular membranes and membrane channels with molecular dynamics. *J Struct Biol*, 157(3):534–544, Mar 2007.
- [179] K. Murata, K. Mitsuoka, T. Walz, P. Agre, J. Heymann, A. Engel, and Y. Fujiyoshi. Structural determinants of water permeation through Aquaporin-1. *Nature*, 407:599–605, 2000.
- [180] B. Ilan, E. Tajkhorshid, K. Schulten, and G. A. Voth. The mechanism of proton exclusion in aquaporin channels. *Proteins*, 55:223–228, 2004.
- [181] J. S. Hub and B. L. de Groot. Does CO<sub>2</sub> permeate through Aquaporin-1? *Biophys J*, 91(3):842–848, Aug 2006.
- [182] Y. Wang, K. Schulten, and E. Tajkhorshid. What makes an aquaporin a glycerol channel? A comparative study of AqpZ and GlpF. *Structure*, 13(8):1107–1118, Aug 2005.
- [183] E. Beitz, B. Wu, L. M. Holm, J. E. Schultz, and T. Zeuthen. Point mutations in the aromatic/arginine region in aquaporin 1 allow passage of urea, glycerol, ammonia, and protons. *Proc. Natl. Acad. Sci. USA*, 103(2):269–274, Jan 2006.
- [184] 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.
- [185] L. J. Smith, H. J. C. Berendsen, and W. F. van Gunsteren. Computer simulation of urea-water mixtures: A test of force field parameters for use in biomolecular simulation. *J. Phys. Chem. B*, 108(3):1065–1071, 2004.
- [186] B. 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. *J. Biol. Chem.*, 272:16140–16146, 1997.



- [187] F. Chaumont, M. Moshelion, and M. J. Daniels. Regulation of plant aquaporin activity. *Biol Cell*, 97(10):749–764, Oct 2005.
- [188] K. Hedfalk, S. Törnroth-Horsefield, M. Nyblom, U. Johanson, P. Kjellbom, and R. Neutze. Aquaporin gating. *Curr Opin Struct Biol*, 16(4):447–456, Aug 2006.
- [189] A. Finkelstein. *Water movement through lipid bilayers, pores, and plasma membranes*. Wiley & Sons, 1987.
- [190] B. L. de Groot, D. P. Tieleman, P. Pohl, and H. Grubmüller. Water permeation through gramicidin A: desformylation and the double helix; a molecular dynamics study. *Biophys. J.*, 82:2934–2942, 2002.
- [191] F. Zhu, E. Tajkhorshid, and K. Schulten. Collective diffusion model for water permeation through microscopic channels. *Phys Rev Lett*, 93(22):224501, Nov 2004.
- [192] A. Engel and H. Stahlberg. Aquaglyceroporins: Channel proteins with a conserved core, multiple functions and variable surfaces. volume 215 of *International Review of Cytology*, pages 75–104. Academic press, New York, 2002.
- [193] G. Chandy, G. A. Zampighi, M. Kreman, and J. E. Hall. Comparison of the water transporting properties of mip and aqp1. *J Membr Biol*, 159(1):29–39, Sep 1997.
- [194] A. N. van Hoek, M. C. Wiener, J. M. Verbavatz, D. Brown, P. H. Lipniunas, R. R. Townsend, and A. S. Verkman. Purification and structure-function analysis of native, PN-Gase F-treated, and endo-beta-galactosidase-treated CHIP28 water channels. *Biochemistry*, 34(7):2212–2219, Feb 1995.
- [195] J. M. Verbavatz, D. Brown, I. Sabolic, G. Valenti, D. A. Siello, A. N. Van Hoek, T. Ma, and A. S. Verkman. Tetrameric assembly of CHIP28 water channels in liposomes and cell-membranes — a freeze-fracture study. *J. Cell Biol.*, 123:605–618, 1993.
- [196] B. Yang, A. N. van Hoek, and A. S. Verkman. Very high single channel water permeability of aquaporin-4 in baculovirus-infected insect cells and liposomes reconstituted with purified aquaporin-4. *Biochemistry*, 36(24):7625–7632, Jun 1997.
- [197] M. L. Zeidel, S. Nielsen, B. L. Smith, S. V. Ambudkar, A. B. Maunsbach, and P. Agre. Ultrastructure, pharmacological inhibition, and transport selectivity of aquaporin channel-forming integral protein in proteoliposomes. *Biochemistry*, 33:1606–1615, 1994.
- [198] B. L. de Groot, A. Engel, and H. Grubmüller. A refined structure of human Aquaporin-1. *FEBS Lett.*, 504:206–211, 2001.
- [199] F. Zhu, E. Tajkhorshid, and K. Schulten. Molecular dynamics study of aquaporin-1 water channel in a lipid bilayer. *FEBS Lett.*, 504:212–218, 2001.

- [200] F. Q. Zhu, E. Tajkhorshid, and K. Schulten. Theory and simulation of water permeation in aquaporin-1. *Biophys. J.*, 86:50–57, 2004.
- [201] S. Nielsen, J. Frøkiaer, D. Marples, T.-H. Kwon, P. Agre, and M. A. Knepper. Aquaporins in the kidney: from molecules to medicine. *Physiol Rev*, 82(1):205–244, Jan 2002.
- [202] L. S. King, M. Choi, P. C. Fernandez, J. P. Cartron, and P. Agre. Defective urinary-concentrating ability due to a complete deficiency of aquaporin-1. *N Engl J Med*, 345(3):175–179, Jul 2001.
- [203] A. S. Verkman. Renal concentrating and diluting function in deficiency of specific aquaporin genes. *Exp Nephrol*, 10(4):235–240, 2002.
- [204] T. Ma, Y. Song, B. Yang, A. Gillespie, E. J. Carlson, C. J. Epstein, and A. S. Verkman. Nephrogenic diabetes insipidus in mice lacking aquaporin-3 water channels. *Proc Natl Acad Sci U S A*, 97(8):4386–4391, Apr 2000.
- [205] P. Agre, S. Nielsen, and O. P. Ottersen. Towards a molecular understanding of water homeostasis in the brain. *Neuroscience*, 129(4):849–850, 2004.
- [206] M. Amiry-Moghaddam, D. S. Frydenlund, and O. P. Ottersen. Anchoring of aquaporin-4 in brain: molecular mechanisms and implications for the physiology and pathophysiology of water transport. *Neuroscience*, 129(4):999–1010, 2004.
- [207] S. Popović-Suić, J. Sikić, N. Vukojević, B. Cerovski, M. Nasić, and R. Pokupec. Target intraocular pressure in the management of glaucoma. *Coll Antropol*, 29 Suppl 1:149–151, 2005.
- [208] S. Steinfeld, E. Cogan, L. S. King, P. Agre, R. Kiss, and C. Delporte. Abnormal distribution of aquaporin-5 water channel protein in salivary glands from Sjögren’s syndrome patients. *Lab Invest*, 81(2):143–148, Feb 2001.
- [209] K. Tsubota, S. Hirai, L. S. King, P. Agre, and N. Ishida. Defective cellular trafficking of lacrimal gland aquaporin-5 in sjögren’s syndrome. *Lancet*, 357(9257):688–689, Mar 2001.
- [210] T. Hibuse, N. Maeda, T. Funahashi, K. Yamamoto, A. Nagasawa, W. Mizunoya, K. Kishida, K. Inoue, H. Kuriyama, T. Nakamura, T. Fushiki, S. Kihara, and I. Shimomura. Aquaporin 7 deficiency is associated with development of obesity through activation of adipose glycerol kinase. *Proc Natl Acad Sci U S A*, 102(31):10993–10998, Aug 2005.
- [211] N. Maeda, T. Funahashi, T. Hibuse, A. Nagasawa, K. Kishida, H. Kuriyama, T. Nakamura, S. Kihara, I. Shimomura, and Y. Matsuzawa. Adaptation to fasting by glycerol transport through aquaporin 7 in adipose tissue. *Proc Natl Acad Sci U S A*, 101(51):17801–17806, Dec 2004.



- [212] C. M. Niemietz and S. D. Tyerman. New potent inhibitors of aquaporins: silver and gold compounds inhibit aquaporins of plant and human origin. *FEBS Lett*, 531:443–447, 2002.
- [213] G. M. Preston, J. S. Jung, W. B. Guggino, and P. Agre. The mercury-sensitive residue at cysteine-189 in the CHIP28 water channel. *J. Biol. Chem.*, 268:17–20, 1993.
- [214] H. L. Brooks, J. W. Regan, and A. J. Yool. Inhibition of aquaporin-1 water permeability by tetraethylammonium: Involvement of the loop E pore region. *Mol. Pharmacol.*, 57:1021–1026, 2000.
- [215] F. J. M. Detmers, B. L. de Groot, E. M. Müller, A. Hinton, I. B. M. Konings, M. Sze, S. L. Flitsch, H. Grubmüller, and P. M. T. Deen. Quaternary ammonium compounds as water channel blockers. *J. Biol. Chem.*, 281:14207–14214, 2006.
- [216] B. X. Yang, J. K. Kim, and A. S. Verkman. Comparative efficacy of  $\text{HgCl}_2$  with candidate aquaporin-1 inhibitors DMSO, gold, TEA(+) and acetazolamide. *FEBS Lett*, 580:6679–6684, 2006.
- [217] A. J. Yool, O. H. Brokl, T. L. Pannabecker, W. H. Dantzler, and W. D. Stamer. Tetraethylammonium block of water flux in aquaporin-1 channels expressed in kidney thin limbs of Henle’s loop and a kidney-derived cell line. *BMC Physiol*, 2:4, 2002.
- [218] E. M. Müller, J. S. Hub, H. Grubmüller, and B. L. de Groot. Is TEA an inhibitor for human aquaporin-1? *Pflügers Archiv*, 2008. accepted.
- [219] E. Fischer. Einfluss der Configuration auf die Wirkung der Enzyme. *Ber. Dt. Chem. Ges.*, 27:2985–2993, 1894.
- [220] E. D. Koshland. Application of a theory of enzyme specificity to protein synthesis. *Proc. Natl. Acad. Sci. USA*, 44:98–104, 1958.
- [221] G. Fragneto, T. Charitat, F. Graner, K. Mecke, L. Perino-Gallice, and E. Bellet-Amalric. A fluid floating bilayer. *Europhysics Lett.*, 53(1):100–106, 2001.
- [222] J. Katsaras and T. Gutberlet. *Lipid Bilayers : Structure and Interactions*. Springer, 2000.
- [223] J. Lemmich, K. Mortensen, J. H. Ipsen, T. Hønger, R. Bauer, and O. G. Mouritsen. Small-angle neutron scattering from multilamellar lipid bilayers: Theory, model, and experiment. *Phys. Rev. E*, 53(5):5169–5180, May 1996.
- [224] Y. K. Levine and M. H. F. Wilkins. Structure of oriented lipid bilayers. *Nature (Lond.)*, 230(11):69–72, 1971.
- [225] V. Luzzati. X-ray diffraction studies of lipid-water systems. In D. Chapman, editor, *Biological Membranes*, pages 71–124. Academic Press, New York, 1968.

- [226] V. A. Parsegian, R. P. Rand, N. L. Fuller, and D. C. Rau. Osmotic stress for the direct measurement of intermolecular forces. In L. Packer, editor, *Methods in Enzymology*, volume 127. Academic Press, New York, 1986.
- [227] W. Pfeiffer, T. Henkel, E. Sackmann, W. Knoll, and D. Richter. Local dynamics of lipid bilayers studied by incoherent quasi-elastic neutron-scattering. *Europhysics Lett.*, 8(2):201–206, 1989.
- [228] B. E. Warren. X-ray diffraction in long chain liquids. *Phys. Rev.*, 44(12):0969–0976, 1933.
- [229] M. C. Wiener and S. H. White. Structure of a fluid dioleoylphosphatidylcholine bilayer determined by joint refinement of x-ray and neutron diffraction data. III. Complete structure. *Biophys J*, 61(2):434–447, Feb 1992.
- [230] D. R. Fattal and A. Ben-Shaul. Mean-field calculations of chain packing and conformational statistics in lipid bilayers: comparison with experiments and molecular dynamics studies. *Biophys J*, 67(3):985–995, Sep 1994.
- [231] S. H. Chen, C. Y. Liao, H. W. Huang, T. M. Weiss, M. C. Bellisent-Funel, and F. Sette. Collective dynamics in fully hydrated phospholipid bilayers studied by inelastic x-ray scattering. *Phys Rev Lett*, 86(4):740–743, Jan 2001.
- [232] T. M. Weiss, P.-J. Chen, H. Sinn, E. E. Alp, S.-H. Chen, and H. W. Huang. Collective chain dynamics in lipid bilayers by inelastic x-ray scattering. *Biophys J*, 84(6):3767–3776, Jun 2003.
- [233] M. C. Rheinstädter, C. Ollinger, G. Fragneto, F. Demmel, and T. Salditt. Collective dynamics of lipid membranes studied by inelastic neutron scattering. *Phys Rev Lett*, 93(10):108107, Sep 2004.
- [234] E. Egberts, S. J. Marrink, and H. J. C. Berendsen. Molecular-dynamics simulation of a phospholipid membrane. *Eur. Biophys. J.*, 22:423–436, 1994.
- [235] H. Heller, M. Schäfer, and K. Schulten. Molecular-dynamics simulation of a bilayer of 200 lipids in the gel and in the liquid-crystal phases. *J. Phys. Chem.*, 97:8343–8360, 1993.
- [236] E. Lindahl and O. Edholm. Mesoscopic undulations and thickness fluctuations in lipid bilayers from molecular dynamics simulations. *Biophys J*, 79(1):426–433, Jul 2000.
- [237] S.-J. Marrink and H. J. C. Berendsen. Simulation of water transport through a lipid-membrane. *J. Phys. Chem.*, 98:4155–4168, 1994.
- [238] S. J. Marrink and H. J. C. Berendsen. Permeation process of small molecules across lipid membranes studied by molecular dynamics simulations. *J. Phys. Chem.*, 100(41):16729–16738, 1996.





- [239] S.-J. Marrink, E. Lindahl, O. Edholm, and A. E. Mark. Simulation of the spontaneous aggregation of phospholipids into bilayers. *J. Am. Chem. Soc.*, 123:8638–8639, 2001.
- [240] M. Tarek, D. J. Tobias, S.-H. Chen, and M. L. Klein. Short wavelength collective dynamics in phospholipid bilayers: A molecular dynamics study. *Phys. Rev. Lett.*, 87(23):238101, Nov 2001.
- [241] D. P. Tieleman and H. J. C. Berendsen. Molecular dynamics simulations of fully hydrated DPPC with different macroscopic boundary conditions and parameters. *J. Chem. Phys.*, 105:4871–4880, 1996.
- [242] A. Spaar and T. Salditt. Short range order of hydrocarbon chains in fluid phospholipid bilayers studied by x-ray diffraction from highly oriented membranes. *Biophys. J.*, 85:1576–1584, 2003.
- [243] T. J. McIntosh and S. A. Simon. Area per molecule and distribution of water in fully hydrated dilauroylphosphatidylethanolamine bilayers. *Biochemistry*, 25(17):4948–4952, Aug 1986.
- [244] J. F. Nagle and S. Tristram-Nagle. Structure of lipid bilayers. *Biochim Biophys Acta*, 1469(3):159–195, Nov 2000.
- [245] J. B. Klauda, N. Kucerka, B. R. Brooks, R. W. Pastor, and J. F. Nagle. Simulation-based methods for interpreting x-ray data from lipid bilayers. *Biophys. J.*, 90:796–2807, 2006.
- [246] S. Paula, A. G. Volkov, A. N. V. Hoek, T. H. Haines, and D. W. Deamer. Permeation of protons, potassium ions, and small polar molecules through phospholipid bilayers as a function of membrane thickness. *Biophys. J.*, 70:339–348, 1996.
- [247] I. M. de Schepper and E. G. D. Cohen. Very-short-wavelength collective modes in fluids. *J. Stat. Phys.*, 27:223–281, 1982.
- [248] I. M. de Schepper, P. Verkerk, A. A. van Well, and L. A. de Graaf. Short-wavelength sound modes in liquid argon. *Phys. Rev. Lett.*, 50(13):974–977, Mar 1983.
- [249] A. A. van Well and L. A. de Graaf LA. Density fluctuations in liquid neon studied by neutron scattering. *Phys. Rev. A*, 32(4):2396–2412, Oct 1985.
- [250] A. A. van Well, P. Verkerk, L. A. de Graaf, J.-B. Suck, and J. R. D. Copley. Density fluctuations in liquid argon: Coherent dynamic structure factor along the 120-K isotherm obtained by neutron scattering. *Phys. Rev. A*, 31(5):3391–3414, May 1985.
- [251] J. S. Hub, T. Salditt, M. C. Rheinstädter, and B. L. de Groot. Short range order and collective dynamics of dmpc bilayers. A comparison between molecular dynamics simulations, x-ray, and neutron scattering experiments. *Biophys J*, 93:3156–3168, Jul 2007.

- [252] U. Mennicke, D. Constantin, and T. Salditt. Structure and interaction potentials in solid-supported lipid membranes studied by x-ray reflectivity at varied osmotic pressure. *Eur Phys J E*, 20(2):221–230, Jun 2006.
- [253] C. Münster. *Antibiotische Peptide in Phospholipiden: Strukturuntersuchungen mit Beugungsmethoden*. PhD thesis, Sektion Physik, Ludwig-Maximilians-Universität München, 2000.
- [254] E. Prince. *International Tables for Crystallography, Vol. C*. Springer, 3rd edition, 2004.
- [255] R. W. Benz, F. Castro-Román, D. J. Tobias, and S. H. White. Experimental validation of molecular dynamics simulations of lipid bilayers: A new approach. *Biophys. J.*, 88:805–817, 2005.
- [256] F. Castro-Román, R. W. Benz, S. H. White, and D. J. Tobias. Investigation of finite system-size effects in molecular dynamics simulations of lipid bilayers. *J Phys Chem B*, 110(47):24157–24164, Nov 2006.
- [257] D. Constantin, U. Mennicke, C. Li, and T. Salditt. Solid-supported lipid multilayers: structure factor and fluctuations. *Eur Phys J E*, 12(2):283–290, Oct 2003.
- [258] C. Münster, J. Lu, B. Bechinger, and T. Salditt. Grazing incidence x-ray diffraction of highly aligned phospholipid membranes containing the antimicrobial peptide magainin 2. *Eur Biophys J*, 28(8):683–688, 2000.
- [259] A. J. Yool and A. M. Weinstein. New roles for old holes: ion channel function in aquaporin-1. *News Physiol Sci*, 17:68–72, Apr 2002.
- [260] J. Yu, A. J. Yool, K. Schulten, and E. Tajkhorshid. Mechanism of gating and ion conductivity of a possible tetrameric pore in aquaporin-1. *Structure*, 14(9):1411–1423, Sep 2006.

---

## LEBENS LAUF VON JOCHEN SEBASTIAN HUB

---

### PERSÖNLICHE DATEN

Name	Jochen Sebastian Hub
Adresse	Düstere-Eichen-Weg 31, 37073 Göttingen
Geburtstag, -ort	1. Februar 1978 in Stuttgart
Nationalität	Deutsch

### AUSBILDUNG

1997	Abitur am Paracelsus-Gymnasium in Stuttgart-Hohenheim Prüfungsfächer Mathematik, Physik, Deutsch, Ethik Gesamtnote 1,0
1997–1998	Zivildienst
1998–2004	Universität Stuttgart Immatrikuliert für den Diplomstudiengang Physik
2000	Vordiplom in Physik. Gesamtnote ‘Sehr gut’
2001–2002	University of Oregon, 2002 MSc in Physik
2004	Diplom in Physik. Gesamtnote ‘Mit Auszeichnung bestanden’
2004–2008	Max-Planck-Institut für biophysikalische Chemie, Göttingen Dissertation mit dem Titel: “Selectivity, Regulation, and Inhibition of Aquaporin Channels. A Molecular Dynamics Study” Betreut von Dr. Bert L. de Groot. Note der Dissertation: ‘summa cum laude’ Note der mündlichen Prüfung: ‘summa cum laude’