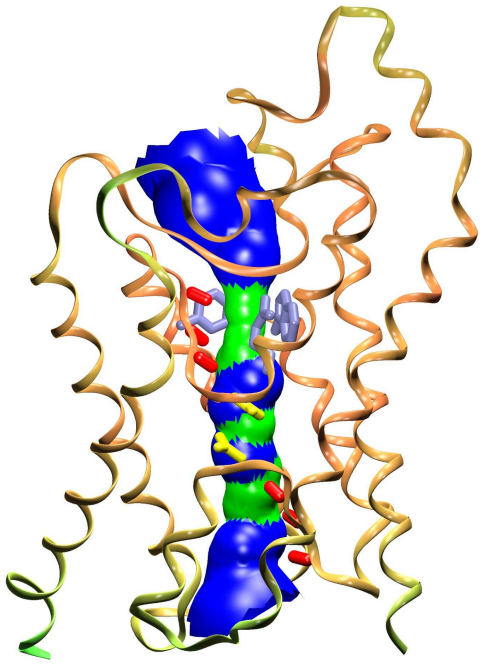
Investigation of Bioenergetical Processes with Multiple-Length-Scale Approaches

Michael Hoffmann



Investigation of Bioenergetical Processes with Multiple-Length-Scale Approaches

Dissertation zur Erlangung des akademischen Grades

Doktor der Naturwissenschaften (Dr. rer. nat.)

vorgelegt dem
Department Physik der Fakultät für Naturwissenschaften
Universität Paderborn

Dipl. Chem. Michael Hoffmann

Paderborn, 2006

Dem Department Physik der Fakultät für Naturwissenschaften als Dissertation vorgelegt.

Eingereicht am:	24.08.2006
Tag der mündlichen Prüfung:	19.12.2006

Promotionskommission

Vorsitzender	Prof. Dr. Gerhard Wortmann
Erstgutachter	Prof. Dr. Thomas Frauenheim
Zweitgutachter	Prof. Dr. Wolf Gero Schmidt
Drittgutachter	Prof. Dr. Marcus Elstner
Beisitzer	Dr. Detlef Schikora

Michael Hoffmann, *Investigation of Bioenergetical Processes with Multiple-Length-Scale Approaches*. PhD Thesis (in English), Department of Physics, Faculty of Science, University of Paderborn, Germany (2006).

Abstract

The study of bioenergetics encompasses energy transformations in organisms. The most common manifestation of this is the ability of organisms to derive energy from its environment, to transform it into a biological useful form, and use it to grow, respond and reproduce. At a molecular and cellular level, a central issue is the coupling of energy-yielding to energy-consuming processes. Simulation of these processes often require the consideration of several length scales ranging from the electronic structure to continuum electrostatics. To handle the computational demands especially involved with conformational sampling, multiple-length-scale approaches that integrate different levels of length scale and theory are an attractive technique. In this work, multiple-length-scale techniques are used to investigate the problem of proton blockage in aquaporins and the phenomenon of color tuning in rhodopsins. Rhodopsins for instance transform light into a proton gradient (bacteriorhodopsin bR) or a photosensory response (phoborhodopsin ppR). The relation of aquaporins to bioenergetics is their to facilitate very efficiently the transmembrane flow of water but preventing dissipation of the energy stored in proton gradients across membranes by impeding proton transfer (PT) through the protein.

For characterizing the progress of long-range PT, first a new reaction coordinate (RC) is proposed and then used to simulate the PT in a model channel. The simulation suggests that this RC can be used efficiently for computing a meaningful potential of mean force. The new RC also eliminates the problems encountered by earlier suggestions and works without assuming a mechanism a priori. In addition, the effect of environments with variable electrostatic properties on the PT energetics demonstrates that the employed QM/MM/continuum electrostatics simulation protocol is capable of describing this aspect of heterogeneous environments found in biological systems.

Using these techniques, the aquaporin GlpF is investigated. The simulated water structure in the pore of GlpF is found to be consistent with previous experimental and theoretical studies. A simulation without proper treatment of long-range electrostatics, in contrast, lacks this pronounced water structure. For the PT through GlpF, a free energy barrier of $\sim$25 kcal/mol, sufficiently high to impede PT through the pore, is found. A perturbation analysis further indicates that the main contribution to the free energy barrier is the desolvation penalty for transferring a proton from bulk solvent to the single water file through the pore rather than distinct structural elements of the protein.

The mechanism of color tuning in the rhodopsin family of proteins is studied by comparing the optical properties of bR and ppR. Despite a high structural similarity, the absorption maximum $\lambda_{\max}$ is strongly shifted between them. Using a coupling of efficient methods, a wide variety of aspects including dynamical effects required for the calculation of absorption spectra are studied. The calculated shift $\Delta\lambda_{\max}$ and the magnitude of the band width agree well with experimental results. Using mutation studies and the analysis of vibrational properties allows the clear identification of two main and equally important factors that are responsible for about 90 % of the spectral shift: the counterion region at the extracellular side of retinal and the amino acid composition of the binding pocket. The good agreement between the theoretical and the experimental results shows that modern quantum mechanical methods can not only reproduce but also interpret spectral properties of photoproteins.

Michael Hoffmann, *Untersuchung von bioenergetischen Prozessen mit Hilfe von Multilängenskalenansätzen*. Dissertation (in englischer Sprache), Department Physik, Fakultät für Naturwissenschaften, Universität Paderborn (2006).

Kurzfassung

Die Bioenergetik befasst sich mit der Erforschung der Energieumwandlungen in Organismen. Sie besitzen die Fähigkeit, aus der Umgebung Energie zu gewinnen, in eine verwertbare Form umzuwandeln und zu Wachstum, Irritabilität und Reproduktion einzusetzen. Ein zentraler Aspekt auf molekularer und zellulärer Ebene ist die Kopplung von energieerzeugenden und energieverbrauchenden Prozessen, deren Simulation oftmals die Betrachtung mehrerer Längenskalen, angefangen von der elektronischen Struktur bis hin zu Kontinuumselektrostatik, erfordert. Um die damit verbundenen Ressourcenanforderungen insbesondere beim "conformational sampling" zu bewältigen, bietet sich die Kopplung von unterschiedlichen Längenskalen und Theorien an. In dieser Arbeit werden Multilängenskalenansätze verwendet, um das Phänomen, dass Aquaporine keine Protonen leiten und das der Verschiebung des Absorptionsmaximums in Rhodopsinen zu untersuchen. Rhodopsine wandeln Licht beispielsweise in einen Protonengradienten (Bakteriorhodopsin bR) oder eine photosensorische Reaktion (Phoborhodopsin ppR) um. Die Bedeutung von Aquaporinen für die Bioenergetik besteht in ihrer Eigenschaft, einen effizienten Wassertransport durch Membranen bei gleichzeitiger Blockade von Protonen zu ermöglichen, wodurch die in Protonengradienten gespeicherte Energie erhalten bleibt.

Zunächst wird eine neue Reaktionskoordinate (RK) für die Beschreibung von langreichweitigem Protonentransfer (PT) eingeführt und für die Simulation des PT in einem Modellkanal verwendet. Die Rechnungen zeigen, dass diese RK eine effiziente Berechnung eines aussagekräftigen Potentiales mittlerer Kraft ermöglicht. Ferner löst die neue RK Probleme vorheriger Ansätze und kommt ohne die Vorgabe eines Mechanismus aus. Darüber hinaus zeigt der Einfluss unterschiedlicher dielektrischer Umgebungen auf die Energetik des PT, dass eine QM/MM/Kontinuumselektrostatik-Kopplung geeignet ist, den elektrostatischen Einfluss heterogener Umgebungen in biologischen Systemen korrekt zu beschreiben.

Diese Techniken werden anschließend auf das Aquaporin GlpF angewandt. Die simulierte Wasserstruktur im Kanal stimmt gut mit vorherigen experimentellen und theoretischen Resultaten überein. Ohne die korrekte Behandlung der langreichweitigen Elektrostatik geht diese Wasserstruktur jedoch verloren. Für den PT durch das Aquaporin GlpF wird eine freie Energiebarriere von $\sim$25 kcal/mol berechnet, was ausreicht, um PT durch den Kanal zu verhindern. Mittels einer Störungsanalyse wird weiterhin gezeigt, dass die Desolvatation eines Protons beim Übergang aus dem Solvens auf die Wasserkette durch den Kanal die Barriere dominiert. Einzelne strukturelle Elemente des Proteins haben auf die Barriere vergleichsweise geringen Einfluss.

Der Mechanismus der spektralen Verschiebung in Rhodopsinen wird anhand des Vergleichs der optischen Eigenschaften von bR und ppR untersucht. Deren Absorptionsmaximum ist trotz ausgeprägter struktureller Ähnlichkeit stark verschoben. Unter Zuhilfenahme einer Kopplung effizienter Methoden werden verschiedene Aspekte, einschließlich dynamischer Effekte, die für die Berechnung von Absorptionsspektren notwendig sind, untersucht. Die berechnete Verschiebung $\Delta\lambda_{\max}$ und Bandenbreite stimmt gut mit experimentellen Werten überein. Ferner erlauben Mutationsstudien und die Analyse von Schwingungen die eindeutige Identifikation von zwei gleichwertigen Faktoren, die für $\sim$90 % der spektralen Verschiebung verantwortlich sind: der Bereich der Gegenionen auf der extrazellulären Seite des Retinals und die Zusammensetzung der Bindungstasche. Die gute Übereinstimmung von theoretischen und experimentellen Resultaten verdeutlicht, dass moderne quantenmechanische Methoden in der Lage sind spektrale Eigenschaften von Photoproteinen nicht nur zu reproduzieren sondern auch zu verstehen.

Contents

Contents **v**

List of Figures viii

List of Tables x

1 Introduction **1**

1.1 Bioenergetics 1

1.2 Rhodopsins and the Problem of Color Tuning 2

1.3 Aquaporins and the Problem of Proton Blockage 5

1.4 The Investigation of Bioenergetical Processes with Multiple-Length-Scale Approaches in this Work 7

1.5 Overview over this Work 8

Fundamentals **11**

2 Theoretical Foundations **11**

2.1 Electronic Structure Methods 11

2.1.1 The quantum mechanical many-body problem 12

2.1.2 Density Functional Theory 13

2.1.3 The Density Functional Theory Based Tight Binding Method . 15

2.1.4 Hartree-Fock Theory 19

2.1.5 The Orthogonalization Model 21

2.2 Configuration Interaction 24

2.3 Molecular Force Field 25

2.3.1 The CHARMM force field 26

2.4 Free Energy Simulations 30

2.4.1 Potential of mean force 30

2.4.2 Umbrella Sampling . . . 32
2.4.3 Weighted Histogram Analysis Method . . . 32

3 Multiple-Length-Scale Approaches 35
3.1 The Quantum Mechanical / Molecular Mechanical Method (QM/MM) 36
3.1.1 Basic Concepts . . . 36
3.1.2 The SCC-DFTB/MM Method . . . 39
3.2 Electrostatic Interactions and Solvation in QM/MM Simulations . . . 40
3.2.1 The Generalized Solvent Boundary Potential . . . 41
3.2.2 Generalized solvent boundary potential in the SCC-DFTB/MM framework . . . 44

Results 47

4 Long-Range Proton Transfer in Biomolecular Channels: Developments and a Model System 47
4.1 Reaction Coordinates for Long-Range Proton Transfer . . . 49
4.1.1 Previous Suggestions . . . 50
4.1.2 A New Reaction Coordinate . . . 51
4.1.3 Comparison of Different Coordinates for a Linear Proton Wire 51
4.2 A Model Proton Channel . . . 53
4.3 Concluding Remarks . . . 58

5 Proton Exclusion in the Glycerol Facilitator GlpF 59
5.1 Structure and Glycerol Conductance of GlpF . . . 59
5.2 Proton Blockage in Aquaporins - Current State of Knowledge . . . 64
5.3 The Mechanism of Proton Exclusion in GlpF . . . 67
5.3.1 Computational Details . . . 68
5.3.2 Water Structure in GlpF . . . 70
5.3.3 Potential of Mean Force Calculations . . . 73
5.4 Discussion . . . 76

6 Color Tuning in Archaeal Rhodopsins: Bacteriorhodopsin vs Phoborhodopsin 81
6.1 Structure and Photocycle of Bacteriorhodopsin . . . 82
6.2 Structure and Photocycle of *pharaonis* Phoborhodopsin . . . 84
6.3 Color Regulation - Current State of Knowledge . . . 87
6.4 Computational Strategy and Methods . . . 90

6.5 Color Tuning between Bacteriorhodopsin and Phoborhodopsin . . . 92
6.5.1 Computational Details . . . 92
6.5.2 Comparison of Structures . . . 94
6.5.3 Excitation Energies and Dipole Moments of bR and ppR . . . 96
6.5.4 Perturbation Analysis . . . 104
6.5.5 Mutation Experiments in ppR . . . 107
6.5.6 Vibrational Analysis . . . 109
6.5.7 Dynamics . . . 110
6.6 Discussion . . . 111

7 Summary **117**

References **121**

Acknowledgements **136**

Colophon **138**

List of Figures

1.1 Absorption spectra for human cone pigments and rhodopsin. 3
1.2 Ribbon representation of the aquaglycoporin GlpF with the pore through the protein depicted as solid surface. 5
1.3 Schematic representation of a multiple-length-scale approach as used in this work. 8

2.1 Formation of bonding and antibonding MOs from two atomic (AO) and two orthogonalized atomic orbitals (OAO). 23
2.2 Relevant intramolecular geometry measures used in molecular force fields: a) bond lengths b) bond angles c) dihedral angles d) out-of-plane angle e) Urey-Bradley. 27

3.1 Overview of different regions involved in multiscale simulations. . . 36
3.2 QM/MM partitioning across a covalent bond showing the MM host atom (MMHA), QM host atom (QMHA) and QM link atom (QML). . 39
3.3 Schematic representation of a biomolecular system using the generalized solvent boundary potential (GSBP). 43

4.1 Water wire models for illustrating different reaction coordinates that describe long-range proton transfers. 52
4.2 Geometric and dielectric parameters for the model channel. 55
4.3 Potential of mean force for proton transfer through a purely nonpolar model channel embedded in different dielectric environments. 57
4.4 Potential of mean force for proton transfer through the model channel with dipoles arranged in the channel. 58

5.1 Structure of the GlpF tetramer in the membrane. 60
5.2 Structure of a GlpF monomer. 60
5.3 Structure of the N-terminal segment of GlpF. 61
5.4 Structural details of the NPA motifs region at the interface of the half-spanning helices M3 and M7. 61
5.5 Structural details of the conduction pathway and the water chain through the pore of GlpF. 62
5.6 The shape of the GlpF pore. 62
5.7 Partitioning of the aquaporin system for QM/MM-GSBP simulations. 68
5.8 (A) Computed relative water distribution in the GlpF channel along the membrane norm (z axis) for the vacuum- and the GSBP-based simulations. (B) Water orientation in the GlpF channel illustrated by the order parameter $P_1(z) = \langle \cos\theta_z \rangle$. 72

5.9 (A) Trace of the location of the excess proton calculated using the mCEC. The color displays the value of the ξ_z coordinate from -12.0 Å (blue) to 16.0 Å (red). (B) PMF for the proton transfer through GlpF. . 73

5.10 Water orientation in the GlpF channel for different reference values of the umbrella sampling biasing potential illustrated by the order parameter $P_1(z) = \langle \cos\theta \rangle$; θ is the angle between the membrane normal and the normalized water dipole vector. 74

5.11 Perturbational correction to the PMF for localized structural elements(A) and global aspects (B). 75

6.1 Structure of the bacteriorhodopsin trimer in the purple membrane. . 83

6.2 Structure of bacteriorhodopsin (PDB code: 1c3w) with selected residues highlighted. 83

6.3 Chemical structure of the retinal chromophore in the all-*trans* configuration. 83

6.4 Detailed view of the retinal Schiff base region of bR (PDB code: 1c3w). 83

6.5 Structure of the ppR/HtrII complex consisting of two molecules of ppR and two molecules of HtrII (PDB code: 1h2s). 86

6.6 Superposition of the structure and retinal of bR and ppR. 86

6.7 Detailed view of the retinal Schiff base region of ppR superposed with that of bR. 86

6.8 Computational strategy for the investigation of the color tuning in rhodopsins. 90

6.9 Hydrogen bonded network of bR and ppR. 94

6.10 Calculated contributions to the excitation energy $\Delta_{S_1-S_0}E$ for individual residues for bR (left) and ppR (right) based on eq 6.2. 101

6.11 Correlation of bond length alternation and excitation energy for bR, ppR and the bR/ppR mutant. 110

6.12 Calculated histograms for the excitation energy of the S_0-S_1 transition for bR and ppR. 111

6.13 Correlation of bond length alternation and excitation energy for bR. . 112

List of Tables

4.1 Values for different reaction coordinates (in Å) for the water wires displayed in Fig. 4.1. 53

4.2 Geometric and dielectric parameters for the model channel shown in Fig. 4.2. 55

4.3 Barrier Heights $W^{\ddagger}$ (in kcal/mol) for the proton transport in the model channel obtained using different dielectric environments. . . . 56

5.1 Perturbations of the GlpF channel studied. [a] 75

6.1 Selected geometrical parameters of the retinal binding site for QM/MM optimized structures of this study, X-ray structures, and from previous theoretical studies. 95

6.2 Vertical excitation energies $\Delta E_{S_1-S_0}$ (in eV). 97

6.3 Dipole moments μ of the chromophore (in Debye) for QM/MM minimized structures. 98

6.4 Vertical excitation energies $\Delta E_{S_1-S_0}$ (in eV) for previous theoretical structures using the OM2/MRCI method. 102

6.5 Selected geometrical parameters of the retinal binding site for QM/MM minimized structures of varying QM zones. 103

6.6 Vertical excitation energies $\Delta E_{S_1-S_0}$ (in eV) for varying QM zones using the OM2 method. 103

6.7 Calculated OM2/MRCI shifts of the vertical excitation energies $\Delta\Delta E_{\text{mutant-wild type}}$ (in eV) and position for glycine-mutants of selected residues. 104

6.8 Calculated OM2/MRCI shifts of the vertical excitation energies $\Delta\Delta E_{\text{mutant-wild type}}$ (in eV) and position for glycine-mutants of polar residues. 106

6.9 Calculated shifts of the vertical excitation energies $\Delta\Delta E_{\text{mutant-wild type}}$ (in eV) of various ppR mutants calculated with OM2/MRCI and SORCI. 108

6.10 Calculated shifts of the N-H stretch vibration and bond length alternation for bR, ppR and the bR/ppR mutant. 109

6.11 Selected geometrical parameters and their fluctuations of the retinal binding site for QM/MM molecular dynamic simulations. 112

Chapter 1

Introduction

1.1 Bioenergetics

It is generally accepted that the cell is the smallest biological unit that is has the ability to derive energy from its environment, transform it into a biologically useful form, and utilize it to drive the familiar activities of organisms. In this context bioenergetics is the research of energy transformations in organisms and comprehends all three mentioned aspects. The incipiencies of bioenergetics as a scientific discipline date from the observation that plants produce oxygen and animals consume oxygen by Priestley and Lavoisier in the 19th century.

At the cellular and molecular level a central issue of bioenergetics can be summarized in the phrase "energy coupling". Organisms clearly need energy for their familiar activities and the required energy is provided by the metabolism, for instance by respiration, fermentation or photosynthesis. However, it was an open issue for a long time how the energy-yielding and energy-consuming processes in the cell are coupled.

One answer to this issue was provided by Lohmann, Fiske, and Subbarow, who discovered adenosine triphosphate (ATP) in 1929 [1, 2], and Lipmann, who developed in 1941 the concept of "phosphate-bond energy" as a principle between energy-generating and energy-utilizing cellular processes [3]. While energy-generating processes are coupled to the synthesis of ATP, the hydrolysis of ATP in turn provides the necessary energy for the performance of work. Being the general currency of energy in the cell, ATP thus plays a similar role as money in economics.

A second milestone was the discovery that organisms additionally avail a second completely different energy currency: namely the energy stored as ion gradients across membranes. This concept of energy coupling using ion currents, the chemiosmotic theory, was first generally expressed by Mitchell in the 1960s [4] and rewarded with the Nobel Prize in 1978.

The connective motif is the "coupling ion" - which is pumped by a transport protein ("ion pump") across the membrane at the expense of some energy source; the generated ion concentration gradient is then used by a second transport protein

(”turbine”) to perform some useful work. This circular flow can be understand as a ”biological” circuit.

The elegance of this concept can be illustrated by the photosynthesis in some Halophilic archaea [5, 6]. The cellular membrane of some Halobacteria contains the photoactive protein bacteriorhodopsin [7] (bR) which acts as a simple light driven proton *pump*. It converts the energy of absorbed light into a proton concentration gradient across the membrane. This gradient, in turn, is used to chemiosmotically drive ATP synthase, which in a *pump-turbine* metaphor thus represents the *turbine*. The same fundamental principle is used in the ATP synthesis by respiration or by photosynthesis.

However, the circular flow of ions through membranes in bioenergetics poses the issue that an efficient coupling of the exergonic and endergonic part of the ”biological” circuit depends on an topologically closed insulating membrane with a low ion permeability. Nonetheless, the membrane must be selectively permeable for cellular needs such as essential nutrients or water. One most interesting group of the highly selective membrane transport proteins that facilitate these exchange processes are aquaporins [8] which facilitate the movement of water across the membrane but are impermeable to ions.

1.2 Rhodopsins and the Problem of Color Tuning

Light is used throughout the biosphere as major energy source for organisms. The photosynthetic reaction centers in plants and bacteria convert it into chemical energy in the form of a proton concentration gradient across the cell membrane. The same goal is achieved by the light-driven proton pump bacteriorhodopsin (bR) in some halophilic archaea.

Beyond its use as energy source, organisms use light to gain information about their environment, for example in the vision process in animals or phototaxis in archaea and bacteria. In many cases, the basis for both processes, the energy and information gathering process, are retinylidene proteins (rhodopsins), the most prominent family among the photoreceptor proteins [9, 10].

The family comprises ion transporting proteins and phototaxis receptors from archaea as well as visual pigments in eukarya, like the cone pigments in the human eye [10]. All rhodopsins share some structural features [10]: The light-absorbing chromophore is retinal (vitamin-A aldehyde), a conjugated polyene chain with a β-ionone ring at the end. The apoprotein (opsin) contains seven transmembrane helices forming an internal binding pocket for the retinal which is bound covalently via a protonated Schiff base linkage. The absorption of a photon leads to the isomerization across one of the double bonds of the polyene chain of the chromophore as primary phototransduction event. The isomerization triggers a photocycle during which the proteins pass through a series of structural alterations that leads to vectorial transport of a proton for bR, for instance, or activates a signaling cascade for visional pigments.

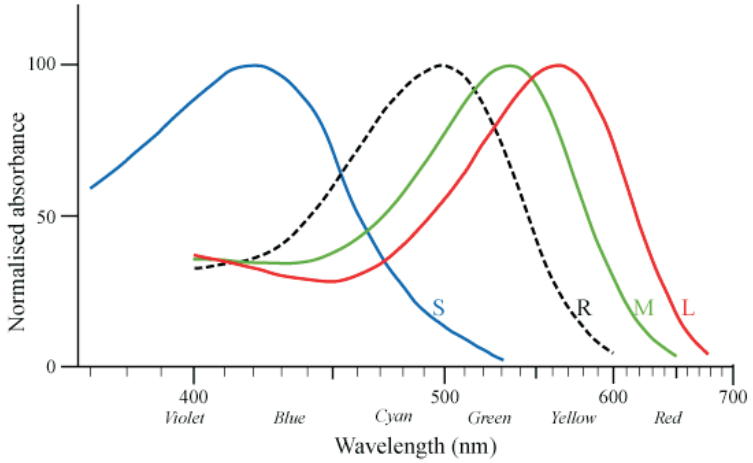


Figure 1.1: Absorption spectra for the human short- (S), medium- (M), and long (L) wavelength sensitive cone pigment and rhodopsin (R). Source: Wikipedia after Bowmaker et al., J. Physiol., **298**, 501 (1980).

The Phenomenon of Color-Tuning

The great pervasiveness of rhodopsins can phenomenologically in part credited to their flexibility to adapt the optical properties of their chromophore to fit the respective biological function. The absorption maximum of the opsin-bound retinal varies over an extreme range from 345 nm (ultraviolet) to 610 nm (red) [11, 12]. This provides organisms with the ability to detect color, on the one hand, and the possibility to adapt the pigments to be most sensitive to the predominating light conditions of the habitat and respective function on the other hand [12–15].

The human eye, for example, contains pigments in the cone cells, often referred to as cone pigments, which are sensitive at short (410 nm, blue), medium (532 nm, green), and long wavelength (563 nm, red) and are responsible for the color discrimination [16] (Fig. 1.1). The pigments in the rod cells, on the other hand, referred to as rhodopsin, are responsible for dim-light vision and absorb at ∼500 nm [17], which is near the maximum of the spectrum of the solar radiation passed through the atmosphere, to attain maximal sensitivity.

For the same reason deep-see fish have pigments that are perfectly adapted to the prevailing blue light of the depths of the oceans. On the other hand, fish living in freshwater often additionally have pigments sensitive to red light, since freshwater is relatively transparent to red light [12, 13]. Some fish even show UV sensitivity as some birds and insects do [12, 14]. While the advantage of UV vision of fish is not evident, many flowers have markings with UV reflectance attracting pollinating insects and birds. Only the fact that the spectral sensitivity of man and insect overlap enables us to catch a glimpse of some of their visual displays not originally intended for us.

The halophilic archaea *Halobacterium salinarum*, on the other hand, lives in nearly saturated salt brine exposed to bright sunlight [18]. It has the ability to switch from

respiratory to photochemical ATP synthesis, if the concentration of oxygen is to low for respiration (semianaerobic conditions) – a flexibility that is rather unique among both simple and complex organisms. Besides bR, its membrane contains three more rhodopsins: halorhodopsin (hR), sensory rhodopsin I (sRI), and phoborhodopsin (pR) [10, 19]. bR and hR carry out light-driven proton and chloride pumping, respectively, across the membrane. This results in a hyperpolarization of the membrane applied to synthesize ATP. The sensory rhodopsins (sRI, pR), on the other hand, are phototaxis receptors that convert light into attractant (sRI) or repellent (pR) migratory behavior of the archaea.

The absorption maximum of the four rhodopsins is partly significantly different reflecting their adaption to the biological function. pR is a blue light-activated repellent receptor with an absorption maximum at $\sim$500 nm (blue-green light) [20], near the maximum of the solar spectrum at the Earth's surface. Its production is induced under aerobic conditions in order to guide the cell to darkness to avoid possible photooxidative damage. In contrast, the absorption maximum of bR, sRI, and hR is red-shifted by about 70 nm relative to pR to 560-590 nm (yellow-orange light) [10]. They are synthesized under semianaerobic conditions and the attractant phototaxis of sRI guides the cell to orange light, where the proton pumping of bR is maximally activated.

Summarizing, there are miscellaneous photic environments throughout the biosphere – ranging from the dim light in the depth of oceans to the bright sunlight in deserts–and rhodopsins have been perfectly adapted to each of these conditions. The wavelength regulation or color tuning problem means therefore at least at a molecular level to identify the mechanisms that determine the wavelength pigments absorb at. The molecular details of these mechanisms are yet a subject of considerable debate.

A protonated retinal Schiff base in methanol absorbs at 440 nm [21]. The shift of the absorption maximum between this value and that in proteins is called 'opsin shift' and must arise from the unique protein-chromophore interaction in each pigment. Recently, the determination of the absorption maximum in vacuum succeeded [22] so that in principle the opsin shift can now defined with respect to the vacuum value. Closely related to the opsin shift is the absorption shift between different proteins, which basically measures the difference between the interaction of two apoproteins with the chromophore.

Numerous experimental studies for sundry rhodopsins have revealed details on the mechanisms determining the opsin shift and the shift between different proteins [21, 23–31]. For example, a highly twisted conformation of retinal has been found in various proteins [32]. The experimental techniques range from comparison of the amino acid sequence of rhodopsins [33–36] over UV-VIS [23, 24, 37, 38], Raman [29, 39–42], and NMR studies [25, 43–47] of retinal or retinal analogs in solution, solid state or various proteins to spectroscopic investigation of mutants [26, 28, 48].

Nevertheless, the mechanistic details are unclear on an atomistic level, wherefore theoretical investigations have been used as an complementary approach and provided further insight [49–62]. But a coherent or quantitative explanation is often difficult to achieve. Computational techniques are in principle able to compute excitation energies for retinal. For the wavelength regulation, however, the protein

environment plays a crucial role as it determines the geometry, dynamics, and electrostatic embedding of the chromophore. Hence, the effect of the protein on the chromophore has to be included. To account for thermal fluctuations of the geometry and electrostatic potential, conformational averaging has to be performed. Therefore, the choice of a suitable method with the required accuracy and the ability to reproduce environmental influences correctly is of particular importance as well as a simulation setup which allows an efficient description of the complete system.

This work will in parts attend to the color tuning in archaeal rhodopsins, in particular the spectral shift between bR and pR. These rhodopsins provide an ideal test case for the investigation of the mechanisms of the color tuning, because of the numerous experimental information available.

1.3 Aquaporins and the Problem of Proton Blockage

Water is the major component of cells. Its movement between cells is mainly restricted by the cell membranes which are highly effective barriers protecting the enclosed compartment. While the membrane itself is impermeable for ionic and most molecular species, embedded transport proteins provide selective permeability for cellular needs to be satisfied. Essential nutrients, specific ions, and unwanted substances are brought in and released, respectively, and water is exchanged with the cells environment in order to maintain its volume.

Among the highly selective membrane transport proteins which are involved in these processes, the facilitation of rapid water permeation across membranes is provided by aquaporin water channels [8]. The family thus plays a key role in establishing and maintaining organism-wide and tissue-specific homeostasis.

Before the discovery of the aquaporins, it had been assumed for a long time that water is exchanged due to simple diffusion through the membrane. This view, however, was challenged by multiple experimental observations (reviewed in refs [63–65]), such as a tissue specific intrinsically high water permeability (e.g., red blood

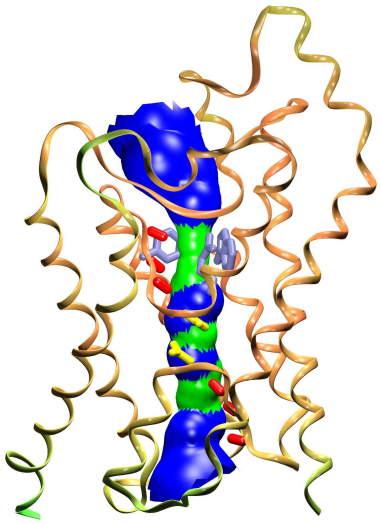


Figure 1.2: Ribbon representation of the aquaglycoporin GlpF with the pore through the protein depicted as solid surface.

cells or renal tubules), which suggested a proteinaceous nature of the water transport. The issue, however, remained disputed until the discovery of the first aquaporin, now designated aquaporin-1 (AQP1), in the late 1980s and early 1990s in the group of Peter Agre [66–69].

Today more than 350 aquaporins are known and members of the family have been found in virtually every organism, ranging from plants, microbials, and invertebrates to vertebrates. In humans alone eleven highly specialized aquaporins (AQP0 to AQP10), which are expressed specific to the respective tissue, are known by now [8, 70, 71].

Their physiological importance is further emphasized by the actuality that their site of expression often coincide with clinical phenotypes: AQP1 and AQP2, for instance, are highly abundant in renal tissue [72–74] and have been confirmed to be involved in abnormal kidney function [75]. Individuals with mutations in the AQP2 sequence suffer from a form of nephrogenic diabetes insipidus [76]. AQP0, on the other hand, was identified in ocular lens fiber cells [77] and inherited defects in the gene encoding of AQP0 [78] were demonstrated to cause congenital cataract affecting small children. Interestingly, the mutation affects a highly conserved residue believed to be important for the protein structure [79].

The aquaporin family comprises the classical aquaporins that conduct only water and aquaglycoporins that additionally conduct glycerol and sometimes other small solutes [80–83]. In *Escherichia coli*, for instance, two aquaporins, a classical aquaporin (AqpZ) and an aquaglycoporin (GlpF), have been found.

The Problem of Proton Blockage

Aquaporins facilitate the transmembrane flow of water, but at the same time the electrochemical properties of the membrane have to be unaffected. Proton concentration gradients across cell membranes or subcellular membranes serve as primary energy source. To maintain these gradients is therefore as important as the water permeation itself.

Aquaporins reconcile these two aspects – high water permeability with an efficient blockage of protons and other ions. This is most intriguing since the proton is the smallest cation and exhibits an outstanding mobility in aqueous solution [84, 85], which is attributed to the fact that it is exceeding rapidly conducted along hydrogen bonded networks of water molecules, known as Grotthuss mechanism [86, 87]. In fact, the mobility of protons is three times higher than that of water itself.

The structure of aquaporins supports a line of water molecules through the protein [88, 88–90] allowing bi-directional single-file passage of water molecules driven by the osmotic pressure close to the diffusion limit. But in contrast to other water filled transmembrane channels such as gramicidin [91–93], aquaporins are impermeable to protons or ions. The reasons for the rejection of ions are without controversy. The energy penalty for conducting ions is enormous [94, 95] since the channels neither are able to accommodate fully hydrated ions because of the pore dimension nor provide residues that could fully solvate ions. The molecular details of the exclusion of protons, however, are highly disputed.

Experimental studies provided the atomic structure of several members of the aquaporin family [82, 96–101] and set the stage for first speculations regarding the mechanism of proton exclusion. Aquaporins are currently among the membrane proteins with the highest atomic resolution. But experimental investigations of the molecular mechanisms remain challenging and only few studies are available [102, 103].

Theoretical analyses, on the other hand, have been proved to provide mechanistic insight at an atomic level regarding both, the permeation of water and the exclusion of protons. Indeed, several studies first mainly addressed the mechanism of the efficient water permeation [88–90, 104, 105] and later increasingly also the mechanism of the water/proton selectivity [88, 104–112]. But despite the effort, the atomic details of the mechanism are still disputed, even if recently the discussion has been reduced to two contradictory explanations.

The controversies reflect the fact that the theoretical studies face a number of challenges. It is indispensable to incorporate the proton transfer (PT) reaction explicitly. The simulation of the PT through aquaporins further involves a large number of groups explicitly. A sufficient conformational sampling of the protein and solvent is thus increasingly inevitable in order to obtain a quantitative estimate of the transfer kinetics. And finally, the electrostatic effect of the protein matrix and highly heterogeneous environment of protein, solvent, and membrane has to be considered adequately.

The adumbrated challenges involve different length scales, wherefore this work will in parts attend to the simulation of the proton exclusion in the aquaglycoporin GlpF using a multiple-length-scale approach.

1.4 The Investigation of Bioenergetical Processes with Multiple-Length-Scale Approaches in this Work

The simulation of the outlined bioenergetical processes involves the consideration of several length scales. The explicit calculation of the proton transfer in aquaporins requires an atomistic description and involves the breaking and formation of bonds. A quantum mechanical treatment or an appropriate approximation, e.g., semiempirical treatment, is therefore required. The same applies to the calculation of excitation energies which requires knowledge about the electronic structure of the chromophore.

For the last years, various quantum mechanical methods have been applied to calculate excitation energies for retinal or proton transfer reactions with increasing accuracy. The potential of these QM methods is enormous considering the speed gain of computational resources in recent years. Nevertheless, with commonly available facilities, the computational costs prohibit the incorporation of the environment, protein or solution, except for model systems or a sufficient amount of conformational sampling at the same level of theory. The incorporation of the protein and solvent environment, however, is crucial for an understanding of the mechanisms involved in the color tuning or the water/proton selectivity.

An intuitive yet powerful alternative to overcome the computational restrictions is to partition the system in a multiple-length-scale approach treating different parts

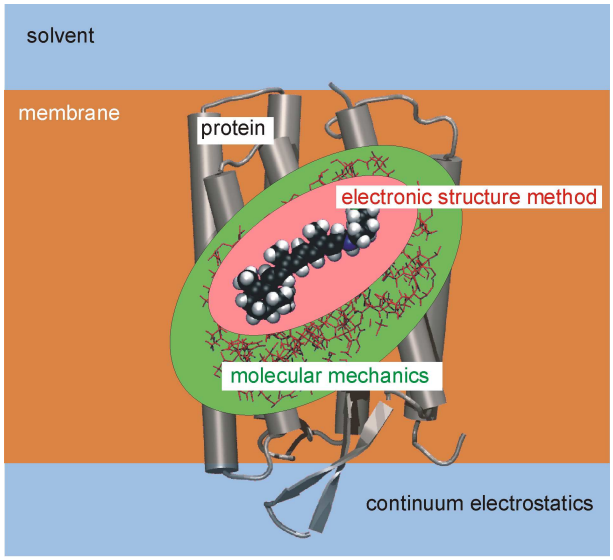


Figure 1.3: Schematic representation of a multiple-length-scale approach as used in this work.

of the system with different levels of theory each with different abilities, accuracies, and computational demands (Fig. 1.3). In this work, the popular hybrid quantum mechanical/molecular mechanical (QM/MM) [113–122] approach combining a quantum mechanical with a molecular mechanical method is used. The quantum mechanical method describes the chemically active region where an explicit description of the electronic structure and the breaking and formation of chemical bonds is required. The molecular mechanical method, on the other hand, allows the incorporation of the remaining protein at an atomic level because of its advantageous accuracy to computational costs ratio.

Furthermore, special care has to be taken to the careful treatment of long-range electrostatics in complex, inhomogeneous protein, solvent, and possible lipid membrane environments, in particular with regard to effects such as the electrostatic shielding due to different dielectric environments [123–131]. In order to consider these effects efficiently, the QM/MM approach is further extended with a coupling to continuum electrostatic methods [128, 130, 132–135] thus incorporating the effect of different dielectric environments. The main advantage of continuum electrostatic method is their great efficiency and the fact that they give the correct long-range behavior for the electrostatic interactions in solution, while the neglect of explicit atomic interactions is their main disadvantage.

The integration of three levels of length-scale and theory - quantum mechanics, molecular mechanics and continuum electrostatics - unites their strengths.

1.5 Overview over this Work

The following two chapters introduce the theoretical fundamentals required for this work. In chapter 2, the methods applied to describe biological systems and their

basis principles are presented. The QM/MM approach and its coupling to a continuum electrostatic approach is introduced chapter 3.

Following, the development of a new reaction coordinate for describing long-range proton transfer is presented in chapter 4. Previous approaches and their drawbacks are discussed. Further, the new reaction coordinate is tested for a model channel using the introduced multiple-length-scale approach.

In chapter 5 and 6, the developed and introduced methods are applied to study the water/proton selectivity in GlpF and the color tuning in archaeal rhodopsins. At the end of this work, a conclusion in chapter 7 summarizes the results.

Chapter 2

Theoretical Foundations

In the following chapter, the methods applied in this work will briefly reviewed. It is subdivided into two topical parts; methods applied to calculate the energy of a molecular system are introduced in the first part, while simulation techniques are outlined in the second one.

Knowing the energy of a molecular system as a function of the nuclear coordinates allows to determine stable conformations, their relative energy, the barriers separating them, or the time dependence of molecular structures and properties. The methods used in this work rely on the atomistic description of molecular systems meaning that the atoms are treated explicitly. They can further separated into methods based on the electronic structure (sec. 2.1) and force field methods (sec. 2.3). Both methods have their advantages and shortcomings wherefore a combination of both in a multiple-length-scale approach is used [136].

For the major part of this work the CHARMM27 force field [137] (sec. 2.3.1) and the SCC-DFTB method [138] (sec. 2.1.3), both implemented in the CHARMM (Chemistry at HARvard Molecular Mechanics) program suite [139], are applied. Their combination has been successfully applied to various biological systems [127, 140–146] and provides an advantageous costs to accuracy ratio. Details about their combination are discussed in the next chapter. The introduction of the OM2 method [147] (sec. 2.1.5) and the MRCI technique [148–150] (sec. 2.2) used to calculate excitation energies complete the first part of this chapter.

In the second part, the fundamentals of free energy computations are shortly outlined. After an introduction to the concept of the potential of mean force, the umbrella sampling technique will be discussed (sec. 2.4.2). Underlying techniques such as molecular dynamics will not be explained. Instead, I refer to corresponding textbooks [149, 151].

2.1 Electronic Structure Methods

The time-dependent Schrödinger equation covers any nonrelativistic problem in the description of the electronic structure of matter. But in most cases, atoms or

molecules without time-dependent interactions are investigated, wherefore one can focus on the time-independent Schrödinger equation:

$$\boldsymbol{H}\psi = E\psi \tag{2.1}$$

2.1.1 The quantum mechanical many-body problem

For an isolated system of N electrons and M nuclei, the Schrödinger equation reads in the stationary case:

$$\widehat{H}_{tot}(\{\boldsymbol{R}\},\{\boldsymbol{r}\})\Psi_{tot}(\{\boldsymbol{R}\},\{\boldsymbol{r}\}) = E\,\Psi_{tot}(\{\boldsymbol{R}\},\{\boldsymbol{r}\}) \tag{2.2}$$

where $\widehat{H}$ is the Hamilton operator and Ψ_{tot} the many-body wave function for the system of electrons and nuclei described by the coordinates of the nuclei $\{\boldsymbol{R}\}$ and electrons $\{\boldsymbol{r}\}$. E denotes the Eigenvalue to $\widehat{H}_{tot}$ and is the total energy of the system.

In the Born-Oppenheimer (BO) approximation the movement of electrons and heavy nuclei is decoupled. For this purpose, the wave function of the system is decomposed into a product of nuclear $\Psi_N(\{\boldsymbol{R}\})$ and electronic $\Psi_{el}(\{\boldsymbol{R}\},\{\boldsymbol{r}\})$ wave functions. The coordinates of nuclei enter into the electronic wave function only parametrically. Thus, the simplified notation is used in the following:

$$\Psi(\{\boldsymbol{r}\}) = \Psi_{el}(\{\boldsymbol{R}\},\{\boldsymbol{r}\}) \tag{2.3}$$

Accordingly, the Hamilton operator of the molecular system is decomposed into an electronic and a nuclear part:

$$\widehat{H}_{tot}(\{\boldsymbol{R}\},\{\boldsymbol{r}\}) = \underbrace{\widehat{T}_N + \widehat{V}_{NN}}_{\widehat{H}_N\{\boldsymbol{R}\}} + \underbrace{\widehat{T}_e + \widehat{V}_{Ne} + \widehat{V}_{ee}}_{\widehat{H}_e(\{\boldsymbol{R}\},\{\boldsymbol{r}\})} \tag{2.4}$$

in which $\widehat{T}$ and $\widehat{V}$ denote the operator of the kinetic and potential energy for the nuclei (N) and electrons (e). The electronic Hamilton operator $\widehat{H}_e$ depends also parametrically on the coordinates of the nuclei via the potential $\widehat{V}_{Ne}$, describing the interaction between nuclei and electrons.

Using the product ansatz for the wave function and Hamilton operator, the Schrödinger equation (eq 2.2) is decomposed into two equations for electrons:

$$\widehat{H}_e\Psi_e(\{\boldsymbol{r}\}) = \left(\widehat{T}_e + \widehat{V}_{Ne} + \widehat{V}_{ee}\right)\Psi_e(\{\boldsymbol{r}\}) = E_e\,\Psi_e(\{\boldsymbol{r}\}) \tag{2.5}$$

and nuclei:

$$\widehat{H}_N\Psi_N(\{\boldsymbol{R}\}) = \left(\widehat{T}_N + \widehat{V}_{NN} + E_e\right)\Psi_N(\{\boldsymbol{R}\}) = E_N\,\Psi_N(\{\boldsymbol{R}\}) \tag{2.6}$$

In this work, the nuclei are treated classically. Therefore, the quantum mechanical problem reduces to finding solutions of the electronic Schrödinger equation (eq 2.5). The subscript "$_e$" will be omitted in the following. The solution yields the electronic states $\{\Psi^m\}$, whereas $m = 0$ refers to the ground state and $m > 0$ to the excited electronic states.

$$\widehat{H}\,\Psi^m(\{\boldsymbol{r}\}) = E^m\,\Psi^m(\{\boldsymbol{r}\}) \tag{2.7}$$

Although the complexity of the many-body problem is reduced significantly, the remaining electronic problem is still a many-body problem of interacting electrons, wherefore an analytically exact solution is impossible. There are several approaches to find approximate solutions. Well-known classes of methods are the Hartree-Fock (HF) method and post-Hartree-Fock methods derived from it [152, 153] and the density functional theory (DFT) [154]. The first and latter are the basis for the OM2 (sec. 2.1.5) and DFTB method (sec. 2.1.3), respectively, used in this work.

2.1.2 Density Functional Theory

The basic idea behind the density functional theory (DFT) of electronic structure is to replace the many-body wave function $\Psi(\{\boldsymbol{r}\})$ by the electron density $n(\boldsymbol{r})$, which depends only on three coordinates instead on $3N$ for the N-electron wave function. In the following, a short introduction into the basic principles is given *.

The Hohenberg-Kohn theorems

The fundamental theorems of the density functional theory by Hohenberg and Kohn [155], legitimize the use of the electron density $n(\boldsymbol{r})$ as basic variable and ensure the existence of a variational principle for the energy.

The first theorem states that *the ground-state electron density determines the external potential.* Thus it determines the number of electrons, the ground-state wave function Ψ, and all other electronic properties of the system including the total energy:

$$E_0 = E[n_0(\boldsymbol{r})] \tag{2.8}$$

However, the exact functional which maps the electron density to the energy is unknown. The universal functional $F[n]$ that maps the electron density n of a free electron gas to its energy $F[n]$ may written within the Hohenberg-Kohn scheme as:

$$F_{HK}[n] = T[n] + V_{ee}[n] \tag{2.9}$$

where $T[n]$ and $V_{ee}[n]$ are the functionals of the kinetic energy and the electron-electron interaction. The expression for the total energy as functional of the electron density then reads:

$$E[n(\boldsymbol{r})] = F_{HK}[n] + \int v_{ext}(\boldsymbol{r})\, n(\boldsymbol{r}) d\boldsymbol{r} = T[n] + V_{ee}[n] + V_{ext}[n] \tag{2.10}$$

The second theorem states the following energy variational principle: *The functional* $E[\tilde{n}] = F[\tilde{n}] + \int v_{ext}(\boldsymbol{r})\, \tilde{n}(\boldsymbol{r}) d\boldsymbol{r}$ *assumes its minimum* E_0 *at the ground state density* $\tilde{n}_0(\boldsymbol{r})$.

The goal of all DFT methods is to find the functional connecting the electron density with the energy. A first approximation by Thomas and Fermi assumed a noninteracting uniform electron gas and explicit approximate forms for $T[n]$ and $V_{ee}[n]$ were constructed. But the calculation of the kinetic energy has constituted a severe problem [154, 156], which was partially solved by Kohn and Sham [157].

*For the rest of this section atomic units will be used. The atomic unit of length is the Bohr radius, the atomic unit of energy is Hartree.

The Kohn-Sham method

The inventive idea by Kohn and Sham [157] was to introduce orbitals in the density functional theory in order to calculate the kinetic energy simply with good accuracy.

The underlying assumption is a system of N noninteracting electrons. It can be described by one-particle wave functions $\psi_i(\boldsymbol{r})$ and associated occupation numbers n_i, with $n_i = 1$ for N orbitals and $n_i = 0$ for the rest. In this case, the ground-state kinetic energy and total electron density can be written as

$$F[n] = T_s[n] = \sum_{i=1}^{N} n_i \left\langle \psi_i(\boldsymbol{r}) \left| \widehat{t} \right| \psi_i(\boldsymbol{r}) \right\rangle = \sum_{i=1}^{N} n_i \left\langle \psi_i(\boldsymbol{r}) \left| -\frac{\Delta}{2} \right| \psi_i(\boldsymbol{r}) \right\rangle \tag{2.11}$$

and

$$n(\boldsymbol{r}) = \sum_{i=1}^{N} n_i |\psi_i(\boldsymbol{r})|^2 \qquad \text{with} \qquad N = \sum_{i=1}^{occ} n_i \tag{2.12}$$

But these expressions are no longer valid for a system of interacting electrons. Then the expression of the kinetic energy in the above form is generally no longer possible. The idea of Kohn and Sham was to assume a reference system of noninteracting electrons for which the ground state electron density is exactly the one of the interacting system.

But the kinetic energy $T_s[n]$ is still not the exact kinetic energy functional $T[n]$. Therefore, a separation of the universal functional $F[n]$ was suggested in such a way that $T_s[n]$ is its exact kinetic energy component:

$$F[n] = T_s[n] + J[n] + E_{xc}[n] \tag{2.13}$$

The coulomb functional $J[n]$ describes the classical electrostatic energy of electron-electron repulsion

$$J[n] = \frac{1}{2} \iint \frac{n(\boldsymbol{r}) n(\boldsymbol{r'})}{|\boldsymbol{r} - \boldsymbol{r'}|} d\boldsymbol{r} d\boldsymbol{r'} \tag{2.14}$$

and the so-called exchange-correlation energy $E_{xc}[n]$ contains the correction for the kinetic energy, the correlation energy as well as the nonclassical part of V_{ee} (eq 2.9), the exchange energy. The exchange-correlation energy is therefore defined as the difference between the exact Hohenberg-Kohn functional $F_{HK}[n]$ and the energy components $T_s[n]$ and $J[n]$

$$E_{xc}[n] = F_{HK} - T_s[n] - J[n] \tag{2.15}$$

The expression for the total energy (eq 2.10) now takes the form:

$$E^{KS} = \sum_{i}^{N} n_i \left\langle \psi_i \left| -\frac{\Delta}{2} \right| \psi_i \right\rangle + \int v_{ext}(\boldsymbol{r}) n(\boldsymbol{r}) d\boldsymbol{r} + J[n] + E_{xc}[n] \tag{2.16}$$

All terms of this expression except the last one are known analytically. But since the Hohenberg-Kohn functional F_{HK} is unknown, $E_{xc}[n]$ is also unknown and must be approximated.

Starting from the expression for the total energy in eq 2.16, the ground state density can be obtained by applying the variational principle which eventually results in an eigenvalue problem, the Kohn-Sham equations, with an effective one-electron Hamilton operator $\widehat{h}$ and the eigenvalues ϵ_i:

$$\widehat{h}\,|\psi_i\rangle = \epsilon_i\,|\psi_i\rangle \qquad i = 1 \ldots N \tag{2.17}$$

$$\widehat{h} = -\frac{\Delta}{2} + v_{eff} = -\frac{\Delta}{2} + v_{ext} + \int \frac{n(\boldsymbol{r'})}{|\boldsymbol{r} - \boldsymbol{r'}|} d\boldsymbol{r'} + v_{xc}[n] \tag{2.18}$$

The exchange-correlation potential $v_{xc}[n]$ is defined as the functional derivative of E_{xc}:

$$v_{xc}[n] = \frac{\delta E_{xc}[n]}{\delta n} \tag{2.19}$$

It is important to note that the operator $\widehat{h}$ acting on the Kohn-Sham orbitals ψ_i depends itself on the density $n[r]$ and hence on the orbitals. Therefore, the set of eqs 2.17 has to be solved self-consistently.

Using the sum over the eigenvalues ϵ_i, the total energy (eq 2.16) can be rewritten as:

$$E_{tot}^{KS} = \sum_i^N n_i \epsilon_i - \frac{1}{2} \iint \frac{n(\boldsymbol{r'}) n(\boldsymbol{r'})}{|\boldsymbol{r} - \boldsymbol{r'}|} d\boldsymbol{r} d\boldsymbol{r'} - \int v_{xc}[n](\boldsymbol{r}) n(\boldsymbol{r}) d\boldsymbol{r} + E_{xc}[n] \tag{2.20}$$

Through the introduction of the N orbitals, the calculation of the kinetic energy is improved. $T_s[n]$, the dominant part of the exact kinetic energy $T[n]$, is handled indirectly but exactly by the equations. The price is that now N Kohn-Sham equations are to be solved, instead of only one in the Thomas-Fermi formulation. Nevertheless, the computational effort is comparable to solving the Hartree equations. But the KS-DFT is exact in principle and capable to fully take into account the correlation of the electrons.

2.1.3 The Density Functional Theory Based Tight Binding Method

Despite the success of the density functional theory over the last years, the calculation of very large systems or the time evolution over long periods is still out of reach [149]. In order to reduce the demands of density functional methods, the density functional theory based tight binding method (SCC-DFTB) was developed [138, 158]. It is related to empirical tight binding schemes [159] but avoids the difficult parametrization by calculating the Hamilton and overlap matrix elements based on fully self-consistent DFT calculations.

SCC-DFTB gains its efficiency from a series of approximations: the self-consistent treatment of DFT is partially discarded, the use of integral approximations, the tabulation of the remaining integrals, and the use of a minimal basis. Nevertheless, SCC-DFTB yields results comparable to the quality of fully self-consistent DFT calculations, wherefore it provides a good accuracy to computational cost ratio [140, 160].

The derivation of SCC-DFTB starts from the expression of the total energy within the Kohn-Sham formulation (eq 2.20) plus the contribution of the nuclear-nuclear coulomb interaction E_{NN}:

$$E_{tot}^{KS} = \sum_{i}^{occ.} n_i \left\langle \psi_i \left| -\frac{\Delta}{2} + v_{ext} + \frac{1}{2}\int \frac{n(\boldsymbol{r'})}{|\boldsymbol{r}-\boldsymbol{r'}|} d\boldsymbol{r'} \right| \psi_i \right\rangle + E_{xc}[n(\boldsymbol{r})] + \underbrace{\frac{1}{2}\sum_{IJ}^{M} \frac{Z_I Z_J}{|\boldsymbol{R_I}-\boldsymbol{r_J}|}}_{E_{NN}} \tag{2.21}$$

Following the ideas of Foulkes and Heydock [161], the electron density $n(\boldsymbol{r})$ is expanded around a reference or input density $n_0(\boldsymbol{r})$:

$$n(\boldsymbol{r}) = \sum_{i}^{N} n_i |\psi_i|^2 = n_0(\boldsymbol{r}) + \delta n(\boldsymbol{r}). \tag{2.22}$$

Then, the total energy is expanded at the reference density $n_0(\boldsymbol{r})$ up to the second order in the density fluctuations δn, so that without the expansion of E_{xc} the total energy reads:

$$\begin{aligned} E_{tot} = &\sum_{i}^{occ.} n_i \left\langle \psi_i \left| -\frac{\Delta}{2} + v_{ext} + \int \frac{n_0(\boldsymbol{r'})}{|\boldsymbol{r}-\boldsymbol{r'}|} d\boldsymbol{r'} + v_{xc}[n_0(\boldsymbol{r})] \right| \psi_i \right\rangle \\ &- \frac{1}{2}\iint \frac{n_0(\boldsymbol{r'})(n_0(\boldsymbol{r}) + \delta n(\boldsymbol{r}))}{|\boldsymbol{r}-\boldsymbol{r'}|} d\boldsymbol{r} d\boldsymbol{r'} \\ &- \int v_{xc}[n(\boldsymbol{r})](\boldsymbol{r})\,(n(\boldsymbol{r}) + \delta n(\boldsymbol{r}))\, d\boldsymbol{r} \\ &+ \frac{1}{2}\iint \frac{\delta n(\boldsymbol{r'})(n_0(\boldsymbol{r}) + \delta n(\boldsymbol{r}))}{|\boldsymbol{r}-\boldsymbol{r'}|} d\boldsymbol{r} d\boldsymbol{r'} + E_{xc}[n(\boldsymbol{r})] + E_{NN} \end{aligned} \tag{2.23}$$

Accordingly, the exchange-correlation functional E_{xc} is expanded into a Taylor series up to the second order in δn. Including this expansion in eq 2.23, one obtains the following equation for the total energy, whereas all contributions containing only linear density fluctuations cancel out:

$$\begin{aligned} E_{tot} = &\sum_{i}^{occ.} n_i \left\langle \psi_i \underbrace{\left| -\frac{\Delta}{2} + v_{ext} + \int \frac{n_0(\boldsymbol{r'})}{|\boldsymbol{r}-\boldsymbol{r'}|} d\boldsymbol{r'} + v_{xc}[n_0(\boldsymbol{r})] \right|}_{\widehat{H}_0[n_0(\boldsymbol{r})]} \psi_i \right\rangle \\ &- \frac{1}{2}\iint \frac{n_0(\boldsymbol{r'})n_0(\boldsymbol{r})}{|\boldsymbol{r}-\boldsymbol{r'}|} d\boldsymbol{r} d\boldsymbol{r'} - \int v_{xc}[n_0](\boldsymbol{r}) n_0(\boldsymbol{r}) d\boldsymbol{r} + E_{xc}[n_0] + E_{nn} \\ &+ \frac{1}{2}\iint \left[\frac{1}{|\boldsymbol{r}-\boldsymbol{r'}|} + \left(\frac{\delta^2 E_{xc}[n]}{\delta n(\boldsymbol{r}) \delta n(\boldsymbol{r'})} \right)_{n=n_0} \right] \delta n(\boldsymbol{r}) \delta n(\boldsymbol{r'}) d\boldsymbol{r} d\boldsymbol{r'} \end{aligned} \tag{2.24}$$

This expression for the total energy resolves into three parts: terms which depend solely on the reference density - the Hamilton operator $\widehat{H}_0$, contributions in the second line including the nuclear-nuclear interaction, and the contributions depending

on the fluctuations of the density δn in the third line. So the total energy can be rewritten as:

$$E_{tot} = \sum_{i}^{occ.} n_i \left\langle \psi_i \left| \widehat{H}_0[n_0(\boldsymbol{r})] \right| \psi_i \right\rangle + E_{rep}[n_0(\boldsymbol{r})] + E_{2nd} \tag{2.25}$$

Without the second order contribution E_{2nd}, the total energy depends only on the reference or input density and eq 2.25 is the basis of a non-self-consistent tight binding scheme (DFTB). It yields reasonable results for systems in which charge transfer can be neglected. The second order correction will be discussed in detail below.

In the repulsive energy contribution E_{rep} are all terms in the second line of eq 2.24 collected:

$$E_{rep}[n_0(\boldsymbol{r})] = -\frac{1}{2} \iint \frac{n_0(\boldsymbol{r'}) n_0(\boldsymbol{r})}{|\boldsymbol{r} - \boldsymbol{r'}|} d\boldsymbol{r} d\boldsymbol{r'} - \int v_{xc}[n_0](\boldsymbol{r}) n_0(\boldsymbol{r}) d\boldsymbol{r} + E_{xc}[n_0] + E_{nn} \tag{2.26}$$

In principle, this contribution to the total energy can be calculated if the reference densities are known. But as common in tight-binding schemes [161], the terms in eq 2.26 are assumed to be pairwise and short-ranged. So E_{rep} can be written as the sum over short-range repulsive pair potentials $U(|\boldsymbol{R_A} - \boldsymbol{R_B}|)$, which depend only on the distance between the nuclei. The justification of the assumptions are outlined in detail elsewhere [162].

$$E_{rep} \approx \frac{1}{2} \sum_{A,B \neq A}^{M} U(|\boldsymbol{R_A} - \boldsymbol{R_B}|) \tag{2.27}$$

The repulsive energy E_{rep} is calculated from the difference between the total energy of a fully self-consistent DFT calculation and the band-structure E_{BS} and second order E_{2nd} term for a reference system for a range of interatomic pair distances $|\boldsymbol{R_A} - \boldsymbol{R_B}|$:

$$E_{rep} = E_{tot}^{KS}(|\boldsymbol{R_A} - \boldsymbol{R_B}|) - E_{BS}(|\boldsymbol{R_A} - \boldsymbol{R_B}|) - E_{2nd}(|\boldsymbol{R_A} - \boldsymbol{R_B}|) \tag{2.28}$$

Hamilton matrix elements and Basis sets

For the calculation of the matrix elements of the Hamilton operator $\widehat{H}_0$, the Kohn-Sham wave function is expanded into a linear combination of atom-centered basis functions (linear combination of atomic orbitals LCAO):

$$|\psi_i(\boldsymbol{r})\rangle = \sum_{\nu} c_{\nu i} \left| \phi_\nu^A(\boldsymbol{r} - \boldsymbol{R}_A) \right\rangle \qquad \nu \in A \tag{2.29}$$

The basis functions $\phi_\nu^A(\boldsymbol{r} - \boldsymbol{R}_A)$ in turn are a linear combination of Slater orbitals determined from a self-consistent DFT calculation using neutral, spherical symmetric pseudo atoms. Furthermore, only a minimal basis is used.

Applying this LCAO ansatz, the matrix elements of the Hamilton matrix transform to :

$$\left\langle \psi_i \left| \widehat{H}_0[n_0(\boldsymbol{r})] \right| \psi_i \right\rangle = \sum_{\nu,\mu} c_{\nu i}^* c_{\mu i} \left\langle \phi_\nu \left| \widehat{H}_0[n_0(\boldsymbol{r})] \right| \phi_\mu \right\rangle = \sum_{\nu,\mu} c_{\nu i}^* c_{\mu i} h_{\mu\nu}^0[n_0(\boldsymbol{r})] \tag{2.30}$$

The further required reference density $n_0(\boldsymbol{r})$ is written as the sum over atomic reference densities:

$$n_0(\boldsymbol{r}) = \sum_A^M n_0^A(\boldsymbol{r}) \tag{2.31}$$

Furthermore, a two center approximation is introduced for the calculation of the matrix elements. Hence, for the calculation of the reference density, only contribution which originate from the both atoms, which ϕ_ν and ϕ_μ are localized on, are taken into account. The matrix elements can then be distinguished into:

$$h^0_{\mu\nu} = \begin{cases} \epsilon_\mu^{\text{free atom}} & \text{if } \mu = \nu \\ \left\langle \phi_\mu \left| \widehat{t} + \widehat{V}[n_0^{A(\mu)}(\boldsymbol{r}) + n_0^{B(\nu)}(\boldsymbol{r})] \right| \phi_\nu^B \right\rangle & \text{if } \nu \neq \mu, A(\mu) \neq B(\nu) \\ 0 & \text{otherwise.} \end{cases} \tag{2.32}$$

In the case that the basis functions are localized on the same atom, the matrix element would be the eigenvalue of the DFT calculation for the pseudo atom. But to ensure the correct dissociation limit the eigenvalue of the free atom $\epsilon_\mu^{\text{free atom}}$ is used. The introduced approximations further implicate the neglect of crystal field ($< \phi_\mu^A | \widehat{V}_0^B | \phi_\mu^A >$) and tree-center terms ($< \phi_\mu^A | \widehat{V}_0^C | \phi_\nu^B >$).

The matrix elements $h^0_{\mu\nu}$ as well as the elements of the overlap matrix $s_{\mu\nu} = \langle \phi_\mu | \phi_\nu \rangle$ are tabulated for a range of nuclear distances.

Second-order corrections

The SCC-DFTB method introduced so far yields reasonable results if the electron density can be well represented as a sum of atomic-like reference densities. But for the simulation of molecules with a significant charge transfer between the atoms the second-order term E_{2nd} has to be taken into account to describe the charge transfer correctly. The extended method is called self consistent charge DFTB (SCC-DFTB) [138].

In line with the procedure in eq 2.31, the density fluctuations $\delta n(\boldsymbol{r})$ are decomposed into atomic contributions $\delta n^A(\boldsymbol{r})$, which are expanded into a series of radial and angular functions. Truncation after the monopole term accounts for the most important contributions and the coefficients c_{lm}^A are identified as fluctuations of so-called Mulliken charges Δq^A.

$$\delta n^A(\boldsymbol{r}) = \sum_{lm} c_{lm}^A F_{lm}^A Y_{lm}^A \approx F_{00}^A Y_{00}^A \Delta q^A \tag{2.33}$$

In an atomic basis with the overlap matrix $\boldsymbol{s}$, the Mulliken charges q^A and the charge fluctuation Δq^A are:

$$q^A = \sum_i^{occ.} n_i \sum_{\mu \in A} \sum_\nu c_{\mu i} c_{\nu i} S_{\mu\nu} \tag{2.34}$$

$$\Delta q^A = q^A - q_0^A \tag{2.35}$$

where q_0^A represents the number of valence electrons of an neutral atom.

Substitution of the density fluctuations in last term of eq 2.24 thus yields the following expression for the second-order contribution:

$$E_{2nd} = \frac{1}{2}\sum_{A}^{M}\sum_{B}^{M}\gamma_{AB}(R_{AB})\Delta q_A \Delta q_B, \qquad \text{with} \tag{2.36}$$

$$\gamma_{AB}(R_{AB}) = \iint\left[\frac{1}{|\boldsymbol{r}-\boldsymbol{r'}|} + \left(\frac{\delta^2 E_{xc}[n(\boldsymbol{r})]}{\delta n(\boldsymbol{r})\delta n(\boldsymbol{r'})}\right)_{n=n_0}\right]\frac{F_{00}^A F_{00}^B}{4\pi}d\boldsymbol{r}d\boldsymbol{r'} \tag{2.37}$$

For a detailed discussion of γ see refs [160, 162].

Total energy and DFTB secular equation

With all energy contributions and approximations the SCC-DFTB total energy expression (eq 2.25) is finally transformed into:

$$E_{tot} = \sum_{i}^{occ.} n_i \sum_{\nu,\mu} c_{\nu i}^* c_{\mu i} h_{\mu\nu}^0 + \frac{1}{2}\sum_{A}^{M}\sum_{B}^{M}\gamma_{AB}(R_{AB})\Delta q_A \Delta q_B + E_{rep} \tag{2.38}$$

The LCAO coefficients which minimize the total energy are found by the variational principle, which leads to a secular problem of the form:

$$\sum_{\nu} c_{\nu i}\left(h_{\mu\nu} - \epsilon_\nu S_{\mu\nu}\right) = 0 \qquad \forall i, \mu \tag{2.39}$$

Since the elements of the Hamilton matrix depend on the Mulliken charges, but these in turn are calculated from the LCAO coefficients $c_{\nu i}$ which are obtained by the solution of the secular equation, a self-consistent treatment is necessary.

2.1.4 Hartree-Fock Theory

The Hartree-Fock (HF) method approximates the exact wave function of the ground state ψ_0 of a N-electron system described by the electronic Hamiltonian $\widehat{H}$ (eq 2.7) by a single Slater determinant constructed from N one-electron wave functions $\chi_i(\boldsymbol{x_i})$. Each of these spin molecular orbitals (MO) only depends on the coordinate of one electron $\boldsymbol{x} = \binom{\boldsymbol{r}}{\omega}$ which is composed of a spatial $\boldsymbol{r}$ and a spin ω variable. A spin orbital $\chi_i(\boldsymbol{x_i})$ is the product of the spatial orbital $\psi_i(\boldsymbol{r})$ and one of the two spin functions $\alpha(\omega)$ or $\beta(\omega)$. The HF approximation is therefore equivalent to the molecular orbital approximation [163]. For convenience a short-hand notation for a normalized Slater determinant is used [152]:

$$\left|\psi_0(\boldsymbol{x_i}, \boldsymbol{x_j}, \cdots, \boldsymbol{x_N}\right\rangle = \left|\chi_i(\boldsymbol{x_i})\chi_j(\boldsymbol{x_j})\cdots\chi_N(\boldsymbol{x_N})\right\rangle \tag{2.40}$$

Applying the variational principle on the electronic energy

$$E_0 = \left\langle \psi_0 \left| -\sum_{i=1}^{N}\frac{\Delta_i}{2} - \sum_{i=1}^{N}\sum_{I=1}^{M}\frac{Z_I}{|\boldsymbol{R}_I - \boldsymbol{r}_i|} + \sum_{i=1}^{N}\sum_{j>1}^{N}\frac{1}{|\boldsymbol{r}_I - \boldsymbol{r}_i|}\right| \psi_0 \right\rangle \tag{2.41}$$

with the variational flexibility in the spin orbitals eventually leads to the canonical Hartree-Fock equations (a detailed derivation can be found elsewhere [152]):

$$\widehat{f}(\boldsymbol{x_i})\,|\chi_i(\boldsymbol{x_i})\rangle = \epsilon_i\,|\chi_i(\boldsymbol{x_i})\rangle \qquad i = 1,\ldots,N \tag{2.42}$$

with the effective one-electron Fock operator $\widehat{f}$

$$\begin{aligned}\widehat{f}(\boldsymbol{x_i}) &= \widehat{t} + \widehat{v} + \widehat{g} = \widehat{h} + \widehat{g} \\ &= -\frac{1}{2}\Delta_i - \sum_I^M \frac{Z_I}{|\boldsymbol{R}_I - \boldsymbol{r}_i|} + \sum_j^N \widehat{J}_j(\boldsymbol{x_i}) - \widehat{K}_j(\boldsymbol{x_i})\end{aligned} \tag{2.43}$$

The complicated many-electron problem (eq 2.7) is thus replaced by an effective one-electron problem with one equations for each spin orbital.

The Fock operator $\widehat{f}$ contains two parts: an one-electron part $\widehat{h}$ including the operator of the kinetic energy and the nuclei-electron interaction and a two-electron part $\widehat{g}$ including the electron-electron interaction expressed by the coulomb- ($\widehat{J}_j(\boldsymbol{x_i})$) and the exchange operator ($\widehat{K}_j(\boldsymbol{x_i})$):

$$\widehat{J}_j(\boldsymbol{x_i})\,|\chi_i(\boldsymbol{x_i})\rangle = \left[\int \chi_j^*(\boldsymbol{x_j}) r_{ij}^{-1} \chi_j(\boldsymbol{x_j}) d\boldsymbol{x_j}\right] |\chi_i(\boldsymbol{x_i})\rangle \tag{2.44}$$

$$\widehat{K}_j(\boldsymbol{x_i})\,|\chi_i(\boldsymbol{x_i})\rangle = \left[\int \chi_j^*(\boldsymbol{x_j}) r_{ij}^{-1} \chi_i(\boldsymbol{x_j}) d\boldsymbol{x_j}\right] |\chi_j(\boldsymbol{x_i})\rangle \tag{2.45}$$

Both operators and thus the Fock operator depend on the solutions $\{\chi_i\}$, wherefore the HF equations are nonlinear and must be solved by iterative procedures.

With restriction to closed-shell systems, the general spin orbitals χ_i can be specified as restricted spin orbitals, i.e., two electrons with opposite spin share the same spatial orbital $\psi_i(\boldsymbol{r})$. For this case, the HF-equations can easily converted to equations depending only on spatial coordinates by integrating out the spin functions:

$$\widehat{f}(\boldsymbol{r_i})\,|\psi_i(\boldsymbol{r_i})\rangle = \epsilon_i\,|\psi_i(\boldsymbol{r_i})\rangle \qquad i = 1,\ldots,N \tag{2.46}$$

with

$$\widehat{f}(\boldsymbol{r_i}) = \widehat{h} + \sum_j^{N/2} 2\widehat{J}_j(\boldsymbol{r_i}) - \widehat{K}_j(\boldsymbol{r_i}) \tag{2.47}$$

In practice, the integro-differential equations (eq 2.46) are not solved numerically – there are no practical schemes for molecules. Instead, the molecular orbitals are expanded into a linear combination of known basis functions $\{\phi_\mu(\boldsymbol{r}) | \mu = 1, 2, \ldots, K\}$, mostly atom centered atomic orbital (AO):

$$\psi_i(\boldsymbol{r_i}) = \sum_\mu^K C_{i\nu}\phi_\mu(\boldsymbol{r}) \tag{2.48}$$

Substituting the expansion into the HF-equations shifts the problem of calculating the HF molecular orbitals to the calculation of the expansions coefficients $C_{i\nu}$ and eventually leads to the Roothaan equations [164], written as matrix equation:

$$\mathbf{FC} = \mathbf{SC}\boldsymbol{\epsilon} \tag{2.49}$$

Beside the matrix of the expansion coefficient **C** and of the molecular orbital energies $\boldsymbol{\epsilon}$, the equation contains the overlap **S** matrix with the elements $S_{\mu\nu} = \langle \phi_\mu | \phi_\nu \rangle$ and the Fock **F** matrix which can be split into a one-electron and a two-electron part:

$$f_{\mu\nu} = h_{\mu\nu} + g_{\mu\nu} \tag{2.50}$$

$$h_{\mu\nu} = t_{\mu\nu} + v_{\mu\nu} = \langle \phi_\mu | \widehat{t} | \phi_\nu \rangle + \langle \phi_\mu | \widehat{v} | \phi_\nu \rangle \tag{2.51}$$

$$g_{\mu\nu} = \sum_{\lambda\sigma} P_{\lambda\sigma} \left[(\mu\nu|\lambda\sigma) - 1/2(\mu\lambda|\nu\sigma) \right] \quad \text{with} \quad P_{\lambda\sigma} = 2 \sum_{j}^{N/2} C_{\lambda j} C_{\sigma j} \tag{2.52}$$

For the two-electron integrals the following notation is used:

$$(\mu\nu|\lambda\sigma) = \int \phi_\mu^*(\boldsymbol{x_i}) \phi_\nu(\boldsymbol{x_i}) r_{ij}^{-1} \phi_\lambda^*(\boldsymbol{x_j}) \phi_\sigma(\boldsymbol{x_j}) d\boldsymbol{r_i} d\boldsymbol{r_j} \tag{2.53}$$

Because of their large number, the calculation and manipulation of the two-electron integrals is the most time consuming part of a HF calculation.

The Roothaan equation (eq 2.49) is a generalized matrix eigenvalue problem, since the overlap matrix is in general not a unit matrix $\mathbf{S} \neq \mathbf{1}$. The reason is the nonorthogonality of the used atomic orbitals. Therefore, in order to transform the Roothaan equation into a standard matrix eigenvalue equation, the basis set is orthogonalized. This leads eventually to the advantageous eigenvalue problem

$$^{\lambda}\mathbf{F}\, ^{\lambda}\mathbf{C} = {}^{\lambda}\mathbf{C}\boldsymbol{\epsilon} \tag{2.54}$$

which can be solved by matrix diagonalization.

The importance of the HF approximations exceeds the fact that it provides an useful method to solve the electronic problem by far. It is also both the starting point for more accurate methods, which include the effect of electron correlation, and for semiempirical methods, which try to reduce the computational costs.

2.1.5 The Orthogonalization Model

As DFT, HF theory is computationally too costly for large systems or extended molecular dynamics simulations. To reduce the costs, approximate molecular orbital theories, often also categorized as semiempirical methods, have been developed. In practical calculations these days, the *neglect of diatomic differential overlap* (NDDO)-based methods MNDO [165], AM1 [166] and PM3 [167] have replaced the traditional methods (CNDO, INDO, NDDO [168, 169]). An extensive overview to the topic can be found in refs [169, 170].

In the following section, the approximations used in semiempirical methods and problems connected with them will be outlined shortly. The focus will be on the NDDO approximation, which is the basis for the Orthogonalization Model 2 (OM2) introduced in the proximate section.

Semiempirical Methods - An Overview

A starting point to introduce semiempirical methods in the framework of HF theory is the Roothaan equation (eq 2.49). But in contrast to the HF theory, semiempirical

methods use the secular equation:

$$^{\text{semi}}\mathbf{F}\ ^{\text{semi}}\mathbf{C} = {}^{\text{semi}}\mathbf{C}\boldsymbol{\epsilon} \tag{2.55}$$

Meaning that the orthogonalization via the transformation $\mathbf{F} \rightarrow {}^{\lambda}\mathbf{F}$ is not explicitly performed, but rather an orthogonal basis is implicitly assumed.

This assumption is closely related to or prerequisite for the use of the *zero differential overlap* (ZDO) or *neglect-of-differential-overlap* (NDDO) approximation [171]. The NDDO approximation assumes that the overlap between two atomic basis functions depending on the same electron coordinates is zero, if the basis functions are located on different atoms:

$$\phi_{\mu}^{A}(\boldsymbol{r}_i)\phi_{\nu}^{B}(\boldsymbol{r}_i) = \delta_{AB} \tag{2.56}$$

Whereas the ZDO approximation neglects any overlap between two atomic basis functions depending on the same electron coordinates:

$$\phi_{\mu}(\boldsymbol{r}_i)\phi_{\nu}(\boldsymbol{r}_i) = \delta_{\mu\nu} \tag{2.57}$$

The implication of both approximations is that all three- and four-center integrals are neglected. In particular the two-electron integrals (eq 2.53) then simplify in the NDDO approximation to:

$$(\mu_A\nu_B|\lambda_C\sigma_D) = \delta_{AB}\delta_{CD}(\mu_A\nu_B|\lambda_C\sigma_D) \tag{2.58}$$

Since the largest part of the time necessary to perform a HF calculation is spent calculating the two-electron integrals (eq 2.53), which are according to amount small, reducing their number thus represents a significant saving of computational costs. However, the neglect of three- and four-center two-electron integrals is only justified if an orthogonal basis is used[†], wherefore semiempirical methods (must) assume an orthogonal basis.

This assumption, however, results due to the neglect of orthogonalization effects in some severe deficiencies of semiempirical methods [175]: (i) the gap between occupied and unoccupied MO is significantly underestimated, wherefore excitations energies are underestimated too; (ii) barriers of internal rotations are underestimated; and (iii) closed-shell repulsion is not included, causing a stable He_2 in CNDO for example.

The reasons are discussed in some detail by Weber [175] and schematically summarized in Fig. 2.1 for the linear combination of two AOs to MOs.

In a calculation based on eq 2.49, the interaction between two nonorthogonal AO leads to an unsymmetric split of the energy levels of the resulting orthogonal MOs. An orthogonalization of the atomic orbitals (AO) to orthogonalized atomic orbitals (OAO) increases their energy relative to their corresponding AOs (Fig. 2.1). But the interaction leads now to a symmetric split, such that the same energy levels of the MOs are obtained.

Semiempirical methods, on the other hand, neglect the nonorthogonality of the AOs due to the use of eq 2.55 and therefore do not account for the destabilization because

[†]Only in an orthogonal basis these integrals are vanishingly small [172–174]

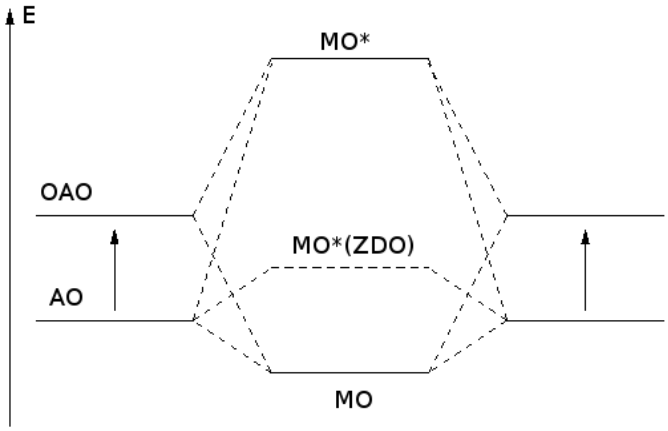


Figure 2.1: Formation of bonding and antibonding MOs from two atomic (AO) and two orthogonalized atomic orbitals (OAO).

of the orthogonalization and yield a symmetric splitting relative to the original AO level. Since the energy of the binding MO relative to the AOs is described correctly due to the parametrization, the energy of antibonding MO is calculated too small.

Further reduction of the computational cost in semiempirical methods is archived by using effective valence shell Hamiltonians - the core electrons are subsumed into the nuclear core - and a minimum basis set. To compensate for introduced errors, empirical parameters are introduced into the remaining integrals and calibrated against reference data.

The Orthogonalization Model 2 - OM2

As outlined above, orthogonalization effects affect the accuracy of semiempirical methods significantly. To overcome the problems, it is necessary to introduce orthogonalization corrections. The most recent approaches are the Orthogonalization Model 2 (OM2) developed by Weber and Thiel [147, 175], and its predecessor the OM1 method by Kolb and Thiel [176].

The OM2 method applies the NDDO integral approximations, a minimal sp-basis set, and includes explicit orthogonalization correction to the one-electron part ($h_{\mu\nu}$) of the Fock matrix elements. The extensions comprise an effective core potential (ECP) which accounts for the orthogonalization effects between core and valence electrons and corrections to the one-center and two-center matrix elements [175] which account for valence-shell orthogonalization effects (ORT). For the two-electron part of the Fock matrix elements, the NDDO approximations are applied without any changes.

In summary, due to the integral approximations and types of interactions included,

the Fock matrix elements characterizing the semiempirical OM2 model are:

$$f_{\mu\mu} = h_{\mu\mu} + \sum_{\nu \text{ on A}} P_{\nu\nu}\left[(\mu\mu|\nu\nu) - 1/2(\mu\nu|\mu\nu)\right] + \sum_{\text{A}\neq\text{B}} \sum_{\lambda \text{ on B}} \sum_{\sigma \text{ on B}} P_{\lambda\sigma}(\mu\mu|\lambda\sigma) \tag{2.59}$$

$$f_{\mu\nu} = h_{\mu\nu} + P_{\mu\nu}\left[\frac{3}{2}(\mu\nu|\mu\nu) - 1/2(\mu\mu|\nu\nu)\right] + \sum_{\text{A}\neq\text{B}} \sum_{\lambda \text{ on B}} \sum_{\sigma \text{ on B}} P_{\lambda\sigma}(\mu\nu|\sigma\lambda) \quad \text{with } \mu,\nu \in A \tag{2.60}$$

$$f_{\mu\nu} = h_{\mu\nu} - \frac{1}{2} \sum_{\lambda \text{ on A}} \sum_{\sigma \text{ on B}} P_{\lambda\sigma}(\mu\lambda|\nu\sigma) \quad \text{with } \mu \in A; \nu \in B \tag{2.61}$$

The corresponding one-electron parts are:

$$h_{\mu\mu} = t_{\mu\mu} + v_{\mu\mu,A} + \sum_{\text{B}\neq\text{A}} \left[v_{\mu\mu,B} + V_{\mu\mu,B}(ECP) + V_{\mu\mu,B}(ORT)\right] \tag{2.62}$$

$$h_{\mu\nu} = \sum_{\text{B}\neq\text{A}} \left[v_{\mu\nu,B} + V_{\mu\nu,B}(ECP) + V_{\mu\nu,B}(ORT)\right] \tag{2.63}$$

$$h_{\mu\nu} = \beta_{\mu\nu} + \sum_{\text{C}} V_{\mu\nu,C}(ORT) \quad \text{with } \beta_{\mu\nu} = t_{\mu\nu} + v_{\mu\nu,A} + v_{\mu\nu,B} \tag{2.64}$$

Regarding the accuracy of the method, Weber [175] could show that both ground state properties, such that enthalpies of formation, ionization energies, and stereospecific properties, and vertical and adiabatic excitation energies[‡] improve significantly compared to MNDO, AM1, and PM3.

2.2 Configuration Interaction

The basic idea of configuration interaction (CI) methods is to represent the exact wave function of the m-th electronic state as a linear combination of N-electron functions (Slater determinants) (eq 2.65).

$$\Phi_0^m = c_0 \Psi_0^{\text{HF}} + \sum_i c_i^m \Psi_i \tag{2.65}$$

The set of N-electron function thereby form the basis of a N-electron function space, the CI space. For the construction of the CI space, a preceding calculation, e.g., a HF calculation, is performed to obtain a set of molecular orbitals (MO). The additional N-electron functions Ψ_i beyond the HF Ψ_i^{HF} are then formed by replacing occupied by unoccupied MOs and usually denoted according to their difference to the HF function. That is functions that differ by n MOs are called n-tuply excited determinants. After generating the CI space, the corresponding energies and coefficients to the trial function eq 2.65 are determined by the linear variation method, i.e., finding the eigenvalues of the matrix representation of the Hamiltonian in the basis of

[‡]The excitation energies were calculated with the PERTCI method [177].

the N-electron functions. The lowest value is then an upper bound to the ground state energy, whereas higher eigenvalues are upper bounds to excited states of the system.

Describing electronic states using a CI expansion (eq 2.65) allows to increasingly consider electron correlation meaning that the averaged electron-electron interaction of the HF approximation is reversed in favor of a more realistic description of the electron correlation (dynamic correlation [178]). On the other hand, it allows the description of multiconfigurational states, i.e., of states that can not correctly described at a single determinant level (static correlation [179]). General reviews regarding electron correlation can be found elsewhere [149, 152, 180].

The different CI methods basically differ in two way: the generation of the CI space and the way the coefficients in front of the N-electron functions are determined. In the full CI (FCI) approach, all possible N-electron functions enter the CI space. However, solving the electronic problem using a FCI approach is a computationally impractical procedure for all but the smallest molecules. Therefore, different strategies have been developed to restrict the CI space, for instance by restricting the excitation level. This simple method reduces the CI expansion in eq 2.65 by considering only those N-electron functions that differ from HF ground state wave function by no more than a predetermined number of MOs. A more complete overview can be found elsewhere [149, 152].

Multi-Reference Configuration Interaction

In the CI methods outlined so far, the CI space is generated by exciting electrons from the HF determinant. In contrast, multi-reference CI (MRCI) methods use a suitable linear combination of N-electron function as first-order space to describe the m-th state. For the case that the number of functions in the expansion is larger than one, the approach is referred to as multi-reference CI approach. The choice of the functions is not unique.

Together with the references, the set of N-electron functions obtained by making all possible excitations on each reference then forms the MRCI space. However, often the number of function in the expansion is too large to be computationally feasible, wherefore the CI expansion on the references is frequently restricted to single and double excitations.

In summary, the major advantages of the MRCI approach are that due to the selection of suitable references a description specific for the selected state is possible. Static correlation is thus mainly covered [148, 150]. Furthermore, major parts of the dynamic correlation are incorporated due to the triple and quadruple excitations included in the full MRCI space. Therefore, the size-consistency error is also reduced [148, 150].

2.3 Molecular Force Field

Besides electronic structure methods, molecular force fields - also referred to as molecular mechanics methods, are a widely used alternative to explore, explain and

predict molecular properties. As electronic structure methods, force fields are based on an atomistic description of the molecular structure, but ignore the electronic motions. Instead the energy of the molecular system is calculated as a parametric function of the nuclear coordinates only. Therefore, force field methods operate within the BO approximation.

Molecular force fields describe the atoms of a molecular system as atom-centered mass points connected by a simple model of the interactions between them including contributions such as the stretching of bonds, the deformation of angles or rotations around bonds. For the atoms, most force field introduce the concept of an atom type. It depends not only on the element of the atom, but it also contains information about the chemical bonding situation the atom is involved in and thus defines how it interacts with other atoms. For instance, the C-C bond length is very similar in different molecules but different from a C=C bond length, although both bonds are between carbon atoms.

The very basis of molecular force fields is the concept that molecules tend to be composed of structural units, which behave similarly in different molecules. This idea leads to the inherent contradiction of force field methods: the compromise between transferability and accuracy.

On the one hand, the force field or more precisely its parametrization should be applicable to a wide variety of molecules. Force fields such as the Universal Force Field (UFF) [181] or the force fields on Allinger et al. [182, 183] head for this.

On the other hand, a force field should be as accurate as possible. So force fields can be reparametrized for the use with specific classes of molecules to improve the accuracy [184]. As example, the CHARMM and AMBER force fields [137, 185] are designed for calculations on proteins and nucleic acids. A detailed overview about available protein force fields can found in a review by Ponder et al. [186].

Furthermore, there is no unique formulation of the functional form, because of the empirical nature of force fields. Therefore another possibility to improve the description is the use of different functional expressions. In principal, many and functional complex terms can be introduced in the force field expression but would foil the transferability in the extreme limit.

2.3.1 The CHARMM force field

The force field CHARMM27 [137] is a class I [187] force field which together with the AMBER [185] and GROMOS [188] force fields are currently used for the majority of biomolecular simulations. The form of the energy functions is similar in these force fields, but there are significant differences in the way the required parameters are determined [187].

The energy expression is written as the sum of contributions each describing the energetic penalty associated with the distortion of the molecule in a specific fashion. The contributions are classified into interactions mediated by bonds, termed intramolecular, and interactions mediated through space, termed intermolecular.

$$E = E_{intra} + E_{inter} \tag{2.66}$$

Intramolecular Interactions - Bonded Interactions

A natural choice of coordinates for the intramolecular interactions are internal coordinates. In internal coordinates, deformations of the molecule are not described with respect to an external Cartesian coordinate system but through the variation of internal bond lengths, angles, and dihedral angles of the molecule (see Fig. 2.2), which are directly derived from the topology. The intramolecular potential then describes the variation of energy connected to deviations from given geometrical parameters (natural parameters). For a discussion of the term natural parameters vs equilibrium parameters see the review by Jensen [149].

The energy function for a deformation of a bond may be written in its simplest form as harmonic approximation of a Taylor expansion around the natural bond length b_0:

$$E_{bond} = \frac{1}{2} k_b (b - b_0)^2 \tag{2.67}$$

where k_b is the associated force constant.

For small deformation from the natural bond length, the harmonic approximation gives sufficient accurate results. However, for larger deviations other functional forms or electronic structure methods have to be applied.

A straightforward improvement is to go beyond the harmonic approximation in the expansion and include cubic and quartic contributions. These anharmonic contributions are for example included in the MM3 and MM4 force fields [182, 183]. Another simple function is the Morse potential [189] which shows the correct asymptotic behavior for bond stretches to infinity.

Similar to the approximations used for bonds, the deformations of bond angles θ (Fig. 2.2b) from their natural bond angle θ_0 can be described as a harmonic potential with the force constant k_θ.

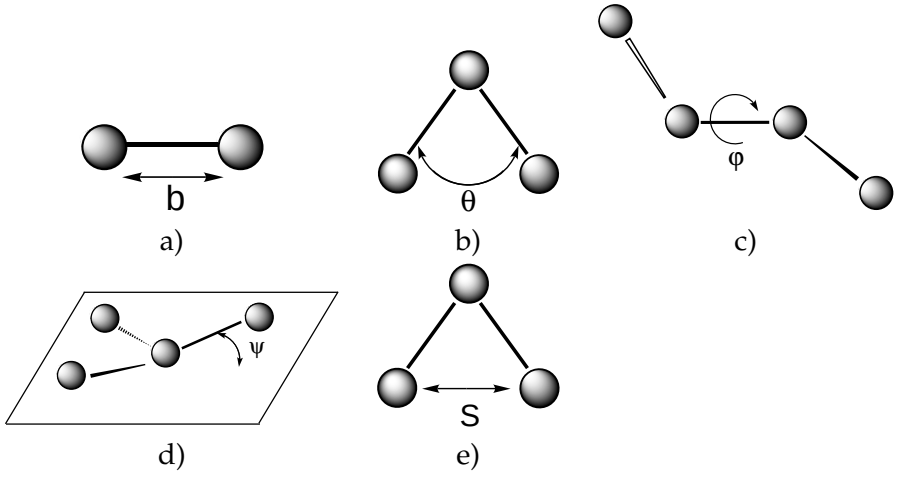


Figure 2.2: Relevant intramolecular geometry measures used in molecular force fields: a) bond lengths b) bond angles c) dihedral angles d) out-of-plane angle e) Urey-Bradley.

For the description of the torsional energy associated with the rotation around the central bond in a four atom sequence (Fig. 2.2c), a different functional form has to be used, since the energy must be periodic in the dihedral angle ϕ and the cost in energy for large deviation from the energetic minimum are often low. A Fourier series with different periodicities n and force constants $k_{n\phi}$ guarantees the requirements.

$$E_{torsion} = \sum_n k_{n\phi} \left(1 + cos(n\phi - \delta_n) \right) \tag{2.68}$$

The use of different periodicities further enables the reproduction of complex rotational energy profiles.

For the special arrangement of three atoms bonded to a central one in the same plane (Fig. 2.2d), for example a carbonyl group, an additional term is necessary to describe deviations from planarity. This out-of-plane contribution can be modelled using a harmonic potential for the angle ψ between a bond from the central atom and the plane defined by the central atom and the remaining two atoms.

In the intramolecular potential of CHARMM, a further contribution is introduced to archive the correct vibrational behavior in some cases, e.g., for H-C-H angles. This Urey-Bradley contribution is defined as harmonic potential for the distance S (Fig. 2.2d) between the outer atoms of a three atom arrangement connected by an angle term.

The expression for the intramolecular contributions to the total energy then reads as follows [§]:

$$\begin{aligned} E_{intra} =& E_{\text{bond}} + E_{\text{angle}} + E_{\text{torsion}} + E_{\text{oop}} + E_{\text{Urey-Bradley}} \qquad (2.69) \\ =& \sum_{bonds} \frac{1}{2} k_b (b - b_0)^2 + \sum_{angles} \frac{1}{2} k_\theta (\theta - \theta_0)^2 + \sum_{torsions} \sum_n k_{n\phi} \left(1 + cos(n\phi - \delta_n) \right) \\ &+ \sum_{impropers} \frac{1}{2} k_\psi (\psi - \psi_0)^2 + \sum_{\text{Urey-Bradley}} \frac{1}{2} k_{UB} (S - S_0)^2 \qquad (2.70) \end{aligned}$$

Intermolecular Interactions - Nonbonded Interactions

Intermolecular interactions or nonbonded interactions include electrostatic and van der Waals interactions, which are mediated through space and not through bonds. For atoms which are separated by less than three bonds [¶] the nonbonded interactions are neglected.

The electrostatic interaction may be modelled as the interaction between fractional atomic charges q_i assigned to the nuclear centers. The interaction between two atoms separated by the distance r_{ij} is then calculated using Coulomb's law with

[§] Note that the indices for looping over the individual geometric measures and parameters have been omitted for clarity

[¶] In some cases nonbonded interactions between atoms separated by three bonds are considered even though scaled down.

ϵ being the electric field constant of the medium:

$$E_{elec} = \frac{1}{4\pi\epsilon} \sum_{\substack{nonbonded,\\ I<J}} \frac{q_I q_J}{r_{IJ}} \tag{2.71}$$

In addition to this simplest approach, there are more elaborated schemes which are discussed in detail in section 3.2 of the next chapter.

The second nonbonded contribution, the van der Waals interactions are described using a Lennard-Jones potential:

$$E_{vdW} = \sum_{\substack{nonbonded,\\ I<J}} 4\epsilon_{IJ} \left[\left(\frac{\sigma_{IJ}}{r_{IJ}} \right)^{12} - \left(\frac{\sigma_{IJ}}{r_{IJ}} \right)^{6} \right] \tag{2.72}$$

where σ is the diameter of collision and ϵ the well depth. For heteroatomic interactions the parameters are computed using the combination rules $\sigma_{ij} = 1/2(\sigma_i + \sigma_j)$ and $\epsilon_{ij} = \sqrt{\epsilon_i \epsilon_j}$.

The van der Waals interactions are zero at long distances, attractive at medium distances, and strongly repulsive for short distances between the atoms. At intermediate distances the van der Waals interaction are modeled using a r^{-6} term. This attraction is due dispersive or London interactions. The dispersive interaction is a result of the electron correlation and can be pictorially understood as a dipole–induced dipole interaction. Fluctuations in the electron density give rise to instantaneous dipoles which in turn can induce dipoles in adjacent electron densities. Contributions from higher order terms such as induced quadrupole–dipole are not considered.

The short–ranged repulsion is modelled using the r^{-12} term and is due to the quantum mechanical exchange repulsion. The functional form of the repulsion is not unique. So a r^{-9} or r^{-10} even exponential dependencies have been used but could not commonly establish because of computational reasons.

Parametrization of Force Fields

Force field parameterization includes the determination of suitable values for the force constants and associated natural parameters of bonds, angles, out-of-plane angles, and Urey-Bradley terms. Further, force constants, periodicities and the phase for the dihedral angles are required. For the description of nonbonded interactions, the partial charges q for the electrostatic energy and the Lennard-Jones Parameters ϵ_{ij} and σ_{ij} for the description of van der Waals interactions have to be determined.

To obtain a reasonable description, all of these parameters have to be adapted for the different atom types. Hence the force field parameters are specific for each term summed over.

Between the different force fields exist significant differences as the parameters are determined. The philosophy behind the CHARMM force field is that the partial charges q_i are determined such that structures and interactions energies are reproduced. It is therefore considered as an interaction based force field. As reference

data for the interaction energies, experimental and ab–initio data are used. For the parametrization the interactions are calculated using the TIP3 water model [190].

In contrast, other force fields assign the partial charges based on electronic structure methods - either on the basis of population analysis methods or via fitting to electrostatic potentials [191–193].

2.4 Free Energy Simulations

The fundamental thermodynamic function of the canonical ensemble is the Helmholtz free energy A. Its calculation is a most important issue, since many physical properties depend on the free energy of the system. However, the calculation of absolute free energies of a complex biomolecular system is utterly impossible. The calculation of the free energy difference between two states, in contrast, is possible and different approaches, for instance thermodynamic integration or thermodynamic perturbation [151], have been suggested. If the change of the free energy is expressed as a function of a spacial coordinate, it is commonly referred to as potential of mean force.

However, the calculation of meaningful potential of mean force is often hampered since the coordinate space is usually not evenly sampled in molecular simulations. This refers to the general problem that ordinary molecular dynamics simulations for protein are limited to the nanosecond time range. Many biological processes such as reactions and conformational transitions, however, take place on a much longer timescale. Also rare activated events or forbidden events, for instance the proton transfer through aquaporins, thus cannot be studied directly. To overcome these obstacles, different techniques have been suggested, such as conformational flooding [194], hyperdynamics [195], and umbrella sampling [196].

After a short introduction to the concept of potential of mean force in the next section, the umbrella sampling technique will be briefly discussed. Subsequently, the weighted histogram analysis method (WHAM) used in this work to calculated the potential of mean force from the umbrella sampling calculations is introduced.

2.4.1 Potential of mean force

The derivation of the potential of mean force starts with a brief overview over distribution functions in classical liquids [197, 198].

Consider a system of N particles with a mass m in a volume V and at a temperature T. The (conservative) forces acting between the particles are described by the interaction potential $\mathcal{U}$. The classical Hamiltonian then reads:

$$H = \sum_i \frac{p_i^2}{2m_i} + \mathcal{U}(\boldsymbol{r}^N) \tag{2.73}$$

where p is the momentum.

The probability of a specific configuration $\{\boldsymbol{r}^N\}$, i.e., that particle number 1 is at $\boldsymbol{r}_1$, particle number 2 at $\boldsymbol{r}_2$, etc., is given by the configurational distribution function

$P^{(N)}(\boldsymbol{r}_1, \boldsymbol{r}_2, ..., \boldsymbol{r}_N)$:

$$P^{(N)}(\boldsymbol{r}_1, \boldsymbol{r}_2, ..., \boldsymbol{r}_N) = \frac{1}{Z} \exp(-\beta \mathcal{U}(\boldsymbol{r}^N)) \tag{2.74}$$

with

$$Z = \int d\boldsymbol{r}^N \exp(-\beta U(\boldsymbol{r}^N)) \qquad \text{and} \qquad \beta = (k_B T)^{-1} \tag{2.75}$$

The probability that particle number 1 is at $\boldsymbol{r}_1$, ..., particle number n at $\boldsymbol{r}_n$, irrespective of the configuration of the remaining $N - n$ particles is obtained by integrating over all coordinates except those of the particles of interest:

$$P^{(n)}(\boldsymbol{r}_1, \boldsymbol{r}_2, ..., \boldsymbol{r}_n) = \frac{1}{Z} \int \dots \int \exp(-\beta \mathcal{U}(\boldsymbol{r}^N)) d\boldsymbol{r}_{n+1} d\boldsymbol{r}_{n+2} \dots d\boldsymbol{r}_N \tag{2.76}$$

In a system, where the particles are indistinguishable, the probability that *any* particle is at $\boldsymbol{r}_1$, ..., and *any* particle is at $\boldsymbol{r}_n$, irrespective of the configuration of the remaining particles, is the reduced distribution function:

$$\rho^{(n)}(\boldsymbol{r}_1, \boldsymbol{r}_2, ..., \boldsymbol{r}_n) = \frac{N!}{(N-n)!} \quad P^{(n)}(\boldsymbol{r}_1, \boldsymbol{r}_2, ..., \boldsymbol{r}_n) \tag{2.77}$$

For $n = 1$ and an isotropic fluid eq 2.77 is equivalent to the particle density ρ:

$$\rho^{(1)}(\boldsymbol{r}_1) = N/V = \rho \tag{2.78}$$

Further, a correlation function $g^{(n)}(\boldsymbol{r}_1, \boldsymbol{r}_2, ..., \boldsymbol{r}_n)$ may be defined:

$$\rho^{(n)}(\boldsymbol{r}_1, \boldsymbol{r}_2, ..., \boldsymbol{r}_n) = \rho^n g^{(n)}(\boldsymbol{r}_1, \boldsymbol{r}_2, ..., \boldsymbol{r}_n) \tag{2.79}$$

$g^{(n)}$ is a called correlation function since it incorporates the correlation between the particles. If they were independent of each other $\rho^{(n)}$ would be simply ρ^n.

$g^{(2)}(\boldsymbol{r}_1, \boldsymbol{r}_2)$ is of particular importance and is termed radial distribution function or pair correlation function. In a fluid of spherically symmetric particles, $g^{(2)}(\boldsymbol{r}_1, \boldsymbol{r}_2)$ depends only upon the relative distance between the particles $r = |\boldsymbol{r}_1 - \boldsymbol{r}_2|$. Therefore, the radial distribution function is simply written as $g(r)$.

In order to describe the interaction of n particles among many other particles in a system, the quantity $w^{(n)}(\boldsymbol{r}_1, \boldsymbol{r}_2, ...\boldsymbol{r}_n)$ is introduced by:

$$g^{(n)}(\boldsymbol{r}_1, \boldsymbol{r}_2, ..., \boldsymbol{r}_n) \equiv e^{-\beta w^{(n)}(\boldsymbol{r}_1, \boldsymbol{r}_2, ...\boldsymbol{r}_n)} \tag{2.80}$$

To see the relevance of this quantity, the average force on a particle 1 for a fixed distance $\boldsymbol{r}_1 - \boldsymbol{r}_2$ to a particle 2 is considered as example:

$$\begin{aligned}
&-\left\langle \frac{d}{d\boldsymbol{r}_1} \mathcal{U} \right\rangle_{\boldsymbol{r}_1, \boldsymbol{r}_2 \text{ fixed}} = \frac{-\int d\boldsymbol{r}_3, ..., d\boldsymbol{r}_N (d\mathcal{U}/d\boldsymbol{r}_1) e^{-\beta \mathcal{U}}}{\int d\boldsymbol{r}_3, ..., d\boldsymbol{r}_N e^{-\beta \mathcal{U}}} = \\
&= k_B T \frac{(d/d\boldsymbol{r}_1) \int d\boldsymbol{r}_3, ..., d\boldsymbol{r}_N e^{-\beta \mathcal{U}}}{\int d\boldsymbol{r}_3, ..., d\boldsymbol{r}_N e^{-\beta \mathcal{U}}} = k_B T \frac{d}{d\boldsymbol{r}_1} \ln \int d\boldsymbol{r}_3, ..., d\boldsymbol{r}_N e^{-\beta \mathcal{U}} \\
&= k_B T \frac{d}{d\boldsymbol{r}_1} \ln \int d\boldsymbol{r}_3, ..., d\boldsymbol{r}_N e^{-\beta \mathcal{U}} + k_B T \underbrace{\frac{d}{d\boldsymbol{r}_1} \ln \frac{N(N-1)}{Z}}_{0} \\
&= k_B T \frac{d}{d\boldsymbol{r}_1} \ln g(\boldsymbol{r}_1, \boldsymbol{r}_2) = -\frac{d}{d\boldsymbol{r}_1} w(r)
\end{aligned} \tag{2.81}$$

The result reveals that the radial distribution function is directly related to the force between two particles averaged over the equilibrium distribution of the remaining particles. The quantity $w(r)$ is therefore the potential whose gradient gives the averaged force bringing the two particles from infinite separation to a separating distance $r = |\boldsymbol{r}_1, \boldsymbol{r}_2|$.

Thus, $w(r)$ is the reversible work and the equation

$$w(\boldsymbol{r}_1, \boldsymbol{r}_2, ..., \boldsymbol{r}_n) = -k_B T \ln g(\boldsymbol{r}_1, \boldsymbol{r}_2, ..., \boldsymbol{r}_n) + c \tag{2.82}$$

is called reversible work theorem. As the derivation of this result suggests, $w(\boldsymbol{r}_1, \boldsymbol{r}_2, ..., \boldsymbol{r}_n)$ is often called the potential of mean force. More generally, for a set of distinguishable particles, the reversible work theorem (eq 2.82) can easily be extended. The change only introduces a constant multiplicative factor to $g(r)$.

2.4.2 Umbrella Sampling

Assume that it is problematic to sample a molecular system described by the potential energy function $\mathcal{U}(\boldsymbol{r}^N)$ efficiently. To overcome the sampling problem, the system is instead described with a modified potential energy function $\mathcal{U}'(\boldsymbol{r}^N)$.

$$\mathcal{U}'(\boldsymbol{r}^N) = \mathcal{U}(\boldsymbol{r}^N) + \mathcal{U}_i(\zeta) \tag{2.83}$$

The artificial biasing potential $\mathcal{U}_i(\zeta)$ is thereby added to the potential energy to extend the range of the sampling of the reaction coordinate ζ which describes the transition of the system between different states. For a simulation including $\mathcal{U}_i$, a histogram of the densities $\rho_i^*(\zeta)$ for the reaction coordinate ζ is generated. The probability density $\rho_i^*(\zeta)$ is then related to the potential of mean force $w_i(\zeta)$ for the nonbiased system by [196]:

$$w_i(\zeta) = -k_B T \ln \rho_i^*(\zeta) - \mathcal{U}_i(\zeta) + c_i \tag{2.84}$$

However, for one particular umbrella potential $\mathcal{U}_i(\zeta)$ the simulation is usually restricted to a part of the reaction coordinate and termed window. In order to cover the complete range of the reaction coordinate, i.e., of the reaction, a large number of windows have to be computed resulting in a set of histograms of the probability density. To obtain the PMF over the entire reaction coordinate from these histograms, the distributions from each of the windows have to be connected. For overlapping distributions, the offset c_i for each window can be obtained by matching the values of $w_i(\zeta)$ at the interfaces between windows. An alternative method is to use the weighted histogram analysis method (WHAM) [199] which is employed in this work and will discussed in the next section. In both cases, a good overlap of the individual histograms is essential to avoid ambiguity at the interfaces.

2.4.3 Weighted Histogram Analysis Method

As introduced above, the potential of mean force is calculated from the probability density $\rho_i^\circ(\zeta)$ of the nonbiased system by

$$w(\zeta) = -k_B T \ln \rho_i^\circ(\zeta) \tag{2.85}$$

Assuming a series of simulations i are conducted with the potential energy function $\mathcal{U}'$ (eq 2.83) where the potential of the nonbiased system is augmented with the umbrella potential $\mathcal{U}_i(\zeta)$. The probability density $\rho_i^\circ(\zeta)$ of the nonbiased system is related to $\rho_i^*(\zeta)$ obtained from simulation i of the perturbed system by

$$\rho_i^*(\zeta) \propto \rho_i^\circ(\zeta) \exp[-\frac{1}{RT}\mathcal{U}_i(\zeta)] \tag{2.86}$$

To determine an estimate $\tilde{\rho}_i^\circ$ for the probability ρ_i°, statistics of the sampling along the reaction coordinate is collected for the simulations. For this purpose, the reaction coordinate is partitioned into bins with the index k and ρ° is replaced by ρ_k°. Subsequently, the number of times $n_{i,k}$ in which the system is found in a particular bin is determined from the simulations and the statistics from all simulations, i.e., the numbers $n_{j,k}$ with $j = 1, 2, ..., i$, are combined. An estimated $\tilde{\rho}_k^\circ$ for the probability ρ_k° of the unperturbed system in a particular bin is then obtained with the WHAM equations, which are derived from eq 2.86.

$$\tilde{\rho}_{j,k}^\circ = \frac{\sum_j n_{j,k}}{\sum_j N_j \tilde{f}_j c_{j,k}} \tag{2.87}$$

and

$$\tilde{f}_j = \frac{1}{\sum_k c_{j,k} \tilde{\rho}_{j,k}^\circ} \tag{2.88}$$

with

$$N_j = \sum_k n_{j,k} \tag{2.89}$$

$$c_{j,k} = \exp[-\frac{1}{RT}\mathcal{U}_j(\zeta_k)] \tag{2.90}$$

In eq 2.87 the summation is over all simulations and in eq 2.88 and 2.89 over all bins. A detailed derivation of the WHAM equations can be found in refs. [199, 200].

The WHAM equations are set of coupled nonlinear equations; the $\tilde{f}_j$ depends on the solution of eq 2.87, wherefore an iterative scheme is used to obtain the $\tilde{\rho}_{j,k}^\circ$ and $\tilde{f}_j$. Starting with an initial estimate for $\tilde{f}_j$, the WAHM equations are iteratively applied until convergence is achieved.

Chapter 3

Multiple-Length-Scale Approaches

The simulation of chemical reactions in biological systems is a formidable challenge. One of the arising problems is the size of systems of real biological interest. A protein typically consists of hundreds of atoms; if the solvent and a possible membrane surrounding it are treated explicitly this number easily exceeds thousands.

Of the methods that are at our disposal, each has its strength and shortcomings. Molecular force fields are certainly able to describe the conformation of biological systems consisting of hundreds of thousands of atoms. But the investigation of the mechanism and energetics of reactions or electronic process are beyond their potential and the domain of electronic structure methods, which are limited to small molecules in return.

One possibility to overcome the restrictions is to integrate a range of methods with different abilities, accuracies, and costs in hybrid or multiscale methods. For this purpose, the system is split up into subsystems each described at a different level of theory (Fig. 3.1). A small section of the system containing the chemically active part is treated by electronic structure methods, while the surrounding protein and solvent molecules are described by molecular force fields. In a further section, continuum electrostatic methods may be used to implicitly incorporate the effects of solvation and a membrane environment.

Considering solvent effects using continuum electrostatics or environmental effects of the protein on an active site attributes to the more general issue of correctly describing long-range electrostatic interactions in complex, inhomogeneous environments, in particular the electrostatic shielding due to different dielectric environments.

In section 3.2, different electrostatic treatments and the generalized solvent boundary potential approach GSBP will be introduced. The coupling of the different methods will be described in section 3.1 and section 3.2.2.

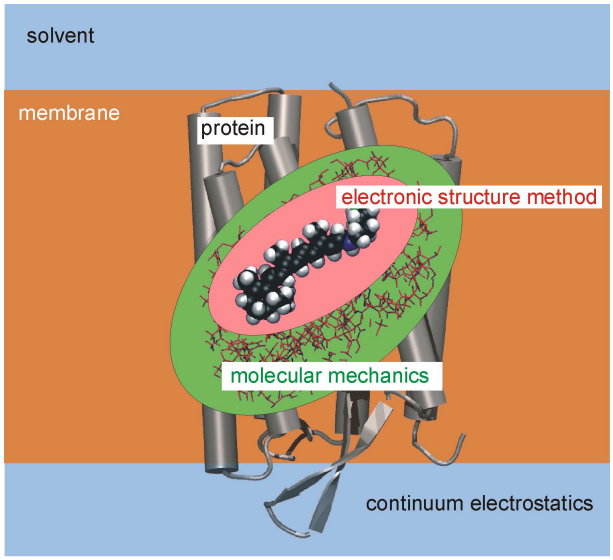


Figure 3.1: Overview of the different regions involved in multiscale simulations used in this work.

3.1 The Quantum Mechanical / Molecular Mechanical Method (QM/MM)

3.1.1 Basic Concepts

Quantum mechanical (QM) approaches can describe the breaking and formation of chemical bonds and hence are required for studying chemical reactivity. But many reactions of interest involve crucial contributions from the environment, which can be solvent or a macromolecule.

However, systems of real biological interest are too large to apply electronic structure methods. Even using semiempirical [166, 201] or approximate density functional theory (DFT) [138] methods, the calculation remains prohibitive at best. To perform sufficient amount of statistical sampling for investigating chemical reactions in complex systems of realistic size faces therefore a severe difficulty. Nevertheless, the potential of these full QM methods is enormous considering the speed gain of computational resources in recent years.

Molecular mechanics (MM) methods, on the other hand, have been widely utilized studying the structural and energetic properties very successively. Their strength comes from their computational efficiency, which predestine them for Molecular Dynamics (MD) or Monte Carlo (MC) simulations. But most force fields are not suitable for problems of chemical reactivity, since they are unable to give an adequate description of the formation and breaking of bonds. Exceptions are reactive force fields, such as the PM6 water force field, which allows the hydrolysis of water [202].

To elude this dilemma, hybrid methods, in which the system is partitioned into several subsystems each treated using a different level of theory, have been developed. The popular hybrid quantum mechanical/molecular mechanical (QM/MM) [113–116]

approach combines the advantages of quantum mechanical methods in describing chemically active regions with a molecular mechanical description of the remaining system.

The underlying assumption is that the changes in electronic structure are localized and that the response of the environment can be described with sufficient accuracy using a force field. Since the introduction of the QM/MM idea around late 70's [116], numerous studies have been conducted examining both performance of different schemes as well as interesting applications to various solution and enzyme systems [117–122]. These previous studies have clearly demonstrated that careful QM/MM methods can provide useful insights into catalytic mechanisms in complex systems that are difficult to obtain otherwise. The current challenges lie in systematically improving the robustness of such methods for general chemical and biochemical systems.

The QM/MM Energy Expression

The available QM/MM schemes can be divided into two groups, an additive and a subtractive scheme, depending as the QM and MM energy contributions are considered.

In the probably most widely used additive scheme, also used in this work, the Hamiltonian operator of the entire system, $\widehat{H}$, is written as the sum of those for the QM partition, $\widehat{H}_{QM}$, the MM partition, $\widehat{H}_{MM}$, and the interaction between the two, $\widehat{H}_{QM/MM}$,

$$\widehat{H} = \widehat{H}_{QM} + \widehat{H}_{MM} + \widehat{H}_{QM/MM} \tag{3.1}$$

The precise expression of $\widehat{H}_{QM/MM}$ varies but generally has contributions from electrostatic, van der Waals, bonded interactions and possibly additional constraints [113],

$$\widehat{H}_{QM/MM} = \widehat{H}^{elec}_{QM/MM} + \widehat{H}^{vdW}_{QM/MM} + \widehat{H}^{bonded}_{QM/MM} + \widehat{H}^{cons}_{QM/MM} \tag{3.2}$$

The bonded terms and constraints (e.g., fixed bond distance between boundary QM and MM atoms) are used for keeping the proper connectivities and geometries when cutting through covalent bonds at the QM/MM interface. The QM/MM van der Waals terms can be optimized to improve properties such as distribution of MM groups around the QM group [203]. The electrostatic component, which is missing in mechanical embedding models (see below) [204], is crucial for polar environments such as water solution and biomolecules (for a recent example, see ref [131]).

The subtractive scheme treats the entire system by MM and an active inner region at the QM and MM level. An appropriate subtraction and addition of the energy contributions eventually yields a suitable correction to the total energy due to the QM method [204].

Coupling Models

Current QM/MM approaches may be grouped into a hierarchy of coupling models based on the way the QM/MM interaction ($\widehat{H}_{QM/MM}$), in particular the electrostatic

interaction ($\widehat{H}^{elec}_{QM/MM}$), is treated [205]:

Mechanical embedding The calculation of the quantum mechanical core is essentially performed without an electrostatic perturbation by the classical region. The coupling is thus limited to steric interactions.

Electrostatic embedding In addition to the mechanical embedding, polarization of the QM region by electrostatic interaction with the MM region is included.

Polarized embedding The polarization of the MM region due to the QM charge distribution is included.

The mechanical embedding scheme has been successfully used in various approaches [204, 206] such as Morokuma's ONIOM model [204]. However, the neglect of polarization of the quantum mechanical region due to the environment is the major deficiency of the scheme. Therefore, the investigation of enzymatic catalysis, for example, which is mainly related to the effect of the environment [126, 131], is extremely problematic.

In the electrostatic embedding scheme, the classical region is incorporated in the quantum mechanical calculation as an external charge distribution. For its description often atomic point charges, either those used in the force field [113, 207] or reparametrized ones if necessary [208], are used. The electrostatic contribution to $\widehat{H}_{QM/MM}$ then reads as follows:

$$\widehat{H}^{QM/MM}_{elec} = \sum_{A \in MM} q_A \int \frac{\rho^{QM}(\boldsymbol{r})}{|\boldsymbol{r} - \boldsymbol{r_A}|} d\boldsymbol{r} \tag{3.3}$$

The polarized embedding scheme additionally allows polarization of the MM region in response to the charge distribution of the QM region thus correcting the asymmetric description of the electrostatic interaction in the previous model. Currently, there are two popular methods to include classical polarization, which are either based on atom-centered polarizable dipoles [116, 209, 210] or on electronegativity equalization [211]. But since the available polarizable force fields are not robust, the electrostatic embedding model is used in this work.

The QM/MM Boundary

An issue that has been repeatedly raised concerns the treatment of the QM/MM boundary. The interaction between MM atoms and nearby QM atoms should be carefully treated to reliably describe the effect of the environment on chemical properties of the QM region. This is expected to be particularly important in cases where the QM/MM partition involves dividing the system across covalent bonds. A typical example involves partitioning catalytic side-chains as QM while the remaining protein as MM. It is evident that special care has to be taken to avoid dangling bonds in the frontier QM atom(s) and to minimize artificial effects on the electronic structure of the QM region.

Various approaches have been proposed to deal with capping of the QM frontier atoms. The most straightforward approach, which is also used in this work, involves

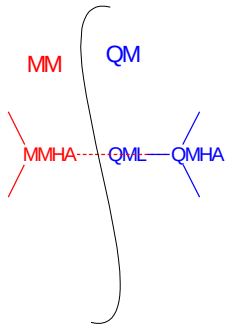


Figure 3.2: QM/MM partitioning across a covalent bond showing the MM host atom (MMHA), QM host atom (QMHA) and QM link atom (QML) [212].

inserting a link atom, which is typically chosen as a hydrogen atom, between the QM host atom (QMHA, see Fig. 3.2) and the MM host atom (MMHA). The link atom is treated at the QM level, and may be subject to an angular and distance constraint to lie along the bond between QMHA and MMHA at a fixed bond distance. The link atom typically interacts with MM atoms through electrostatic terms but not through van der Waals terms. An extensive discussion of the different approaches to treat the electrostatic interaction at the boundary can be found in a recent publication [212]. It contains also a detailed benchmark of the divided frontier charge (DIV) scheme used in this work.

Alternative approaches use hydrogen-like atoms or pseudohalogens instead of regular hydrogen atoms to terminate the QM region [213, 214] or localized orbitals such as the localized self consistent field (LSCF) [215–218] and the generalized hybrid orbital method (GHO) [219] method.

3.1.2 The SCC-DFTB/MM Method

As described in eq 3.1, the potential energy of the QM/MM system is written as

$$E_{tot} = \left\langle \Psi \left| \widehat{H}_{SCC-DFTB} + \widehat{H}^{elec}_{QM/MM} \right| \Psi \right\rangle + U^{vdW}_{QM/MM} + U^{bonded}_{QM/MM} + U_{MM} + U^{cons}_{QM/MM} \tag{3.4}$$

where Ψ is the electronic wave function of the QM region. In this work, the QM level is the SCC-DFTB method [138]. With different frontier treatment schemes [212], the operator describing QM/MM electrostatic interactions, $\widehat{H}^{elec}_{QM/MM}$, takes different forms.

In the original implementation [136], $\widehat{H}^{elec}_{QM/MM}$ has the form of Coulomb interaction between the MM point charge Q_A and the Mulliken charge Δq_B on the QM atom [136].

$$\widehat{H}^{elec,pointcharges}_{QM/MM} = \sum_{A \in MM} \sum_{B \in QM} \frac{Q_A \, \Delta q_B}{|\boldsymbol{r}_A - \boldsymbol{r}_B|} \tag{3.5}$$

The corresponding contribution to the Hamiltonian matrix elements is:

$$\mathbf{H}_{\mu\nu}^{elec,pointcharges} = \frac{1}{2} \sum_{A \in MM} \left(\frac{Q_A}{|\boldsymbol{r}_C - \boldsymbol{r}_A|} + \frac{Q_A}{|\boldsymbol{r}_D - \boldsymbol{r}_A|} \right) \quad \mu \in C, \nu \in D \tag{3.6}$$

3.2 Electrostatic Interactions and Solvation in QM/MM Simulations

Electrostatic interactions are critical to the structure and function of biomolecules [123–130] and challenging to model accurately. An appropriate treatment is most serious for the simulation of processes that involve significant change in the charge distribution, such as oxidation-reduction reactions or long-range proton transfers. An inappropriate treatment of electrostatics may not only affect the quantitative aspect of the result but may also produce qualitative changes in the behavior of biological systems. For example, the experimental water structure in the pore of aquaporins was shown to be only reproduced if a suitable treatment of electrostatic interactions is used (chapter 5).

In contrast to van der Waals interactions which can be smoothly truncated beyond a certain interatomic separation, electrostatic interactions are long-ranged which gives rise to some notable consequences. On the one hand, the calculation of the potential energy in molecular force field simulations is dominated by the costs of calculating the electrostatic interactions. Furthermore, the long-range nature implicates that the incorporation of a solvent environment* becomes an integral part of the problem of evaluating the electrostatic interactions.

The most detailed approach to include solvent environment provide all-atom representations describing the complete biomolecule/solvent system as a collection of atoms interacting with each other via a classical force field. First suggestions to treat the electrostatic interactions in such a system were truncation schemes, in which electrostatic interactions beyond a certain cutoff distance are neglected. These approaches are in particular interesting from the perspective of saving computational costs. However, abrupt truncation schemes lead to instabilities and artifacts in simulations. Therefore, a variety of shifting and switching schemes that modify the Coulomb potential to smoothly truncate have been developed and lead to stable dynamics [220].

Nevertheless, while the use of truncations schemes can be well motivated for short-range van der Waals interactions, it was shown later that they lead to severe artifacts in some simulations [221, 222], especially if long-range electrostatic interactions play an important role. In order to overcome this problem, the extended electrostatics model [223] that combines a standard pairwise additive scheme with a multipole approximation for the calculation of spatially distant interactions was developed.

The most robust electrostatic model using explicit solvent so far is the Ewald summation (or related numerical improvements) [224, 225] with periodic boundary conditions. However, to avoid artifacts in such simulations, a large number of explicit

*The term solvent environment refers here more generally to the complex, inhomogeneous environment around proteins consisting of solvent with or without ions and a possible membrane.

solvent molecules need to be included which limits its use in QM/MM simulations of large systems [226–228]. Hence, it is desirable to adopt schemes that avoid periodic boundary condition.

In this regard, charge scaling schemes [229] and implicit solvent models offer an alternative and have been widely used [124, 125, 133, 230, 231]. While the first ones scale down selected charges of solvent exposed charged residues according to Poisson-Boltzmann calculations to mimic the effect of bulk solvent, the latter use a continuum treatment of solvation.

In macroscopic continuum methods, the macromolecule is treated as macroscopic body of low dielectric constant with embedded fixed charges surrounded by a featureless continuum representation of the solvent. The main advantage of such approaches is their great efficiency, while the neglect of explicit atomic interactions between protein and solvent are their main disadvantage. Also it is believed that they give the correct long-range behavior for the electrostatic interactions in solution.

As a combination of a implicit solvent model with a limited number of explicit solvent molecules, implicit/explicit mixed schemes share much of the advantages of continuum models while adding explicit atomic solvent–solute interactions [132, 133]. The spherical solvent boundary potential (SSBP) model, developed by Roux and coworkers [134], includes a small number of solvent molecules (the first solvation shell) explicitly, while the influence of the remaining bulk solvent is represented with an effective solvent boundary potential.

In the course of allowing accurate simulations of a small region part of a large macromolecule/solvent system, the SSBP methodology was extended to the generalized solvent boundary potential (GSBP) method [135], which is able to incorporate both the influence of macromolecular charge distributions outside the simulation region and the complex electrostatic response from irregular dielectric interfaces. It treats a relatively small region of the protein-solvent-membrane system in microscopic details, with the contribution from the remaining degrees of freedom described at the continuum electrostatics level. This allows very efficient sampling of the most relevant configuration space for the reaction of interest, without significantly sacrificing important environmental effects.

The implementation of GSBP into a QM/MM framework and SCC-DFTB was described in a recent publication [128]. In the following, the GSBP approach in its original form [135] and the implementation in the QM/MM framework are briefly reviewed.

3.2.1 The Generalized Solvent Boundary Potential

In a system composed of biomolecules in solution and membrane, interactions between charge distributions are shielded in a nontrivial way due to complex reaction fields arising from the irregular boundaries between different dielectric environments.

To take into account these effects quantitatively, the electrostatic solvation free en-

ergy has to be considered:

$$\Delta W_{elec}(\boldsymbol{r}) = \frac{1}{2}\sum_{A} q_A \phi_{rf}(\boldsymbol{r}_A) \tag{3.7}$$

with $\phi_{rf}(\boldsymbol{r})$ as the reaction field potential at the position of the atomic charge q_A. The reaction field potential is calculated by subtracting a reference electrostatic potential computed in vacuum ($\phi_v(\boldsymbol{r})$), from the electrostatic potential computed in the dielectric solvent environment ($\phi_s(\boldsymbol{r})$):

$$\phi_{rf}(\boldsymbol{r}) = \phi_s(\boldsymbol{r}) - \phi_v(\boldsymbol{r}) \tag{3.8}$$

The required electrostatic potentials $\phi(\boldsymbol{r})$ are obtained by solving the Poisson-Boltzmann equation:

$$\nabla\left[\epsilon(\boldsymbol{r})\nabla\phi(\boldsymbol{r})\right] - \kappa(\boldsymbol{r})\phi(\boldsymbol{r}) = -4\pi\rho(\boldsymbol{r}) \tag{3.9}$$

where $\rho(\boldsymbol{r})$ is the charge distribution of the explicit atoms using $\rho(\boldsymbol{r}) = \sum_A q_A \delta(\boldsymbol{r} - \boldsymbol{r_A})$, $\epsilon(\boldsymbol{r})$ is the space-dependent dielectric constant, and $\bar{\kappa}(\boldsymbol{r})$ is the space-dependent ion screening length.

Basically, the Poisson-Boltzmann equation (eq 3.9) can be solved numerically for any shape of dielectric boundary and charge distribution, but a repeated solution during a MD simulations is prohibitively expensive for macromolecules. To make the method computational efficient, the generalized solvent boundary potential (GSBP) method [135] was developed.

In the GSBP approach, the system is partitioned into an inner region and an outer environment, where the dielectric property can vary (e.g., containing both bulk solvent and a slab of membrane). Atoms in the inner region are allowed to move during the simulation whereas atoms in the outer region are fixed (Fig. 3.3).

Accordingly, the electrostatic solvation free energy ΔW_{elec} can be partitioned into contributions from the interactions of the outer-outer (oo), inner-outer (io), and inner-inner (ii) parts of the system:

$$\Delta W_{elec} = \Delta W_{elec}^{(oo)} + \Delta W_{elec}^{(io)} + \Delta W_{elec}^{(ii)} \tag{3.10}$$

The outer-outer contribution is constant during a simulation and therefore not of interest. The contribution of the inner-outer part can be computed in a straightforward manner according to eq 3.7,

$$\Delta W_{elec}^{(io)}(\boldsymbol{r}) = \sum_{\alpha \in \text{inner}} q_\alpha \phi_{rf}^{(o)}(\boldsymbol{r}_\alpha) \tag{3.11}$$

where $\phi_{rf}^{o}(\boldsymbol{r}_\alpha)$ is the reaction field due to the outer region atoms, which has to be computed only once (since the outer region is fixed) and saved on a set of grid points in the inner region. Note that actually the *total* electrostatic potential due to the outer region atoms $\phi_s^{(o)}$ is saved instead of the reaction field potential $\phi_{rf}^{(o)}$ so that direct Coulombic interaction between the inner and outer region atoms is included.

The last term arises from the reaction field interactions of the charges in the inner region. But in contrast to the former contributions, the value of $\phi_{rf}^{(i)}$ depends on the

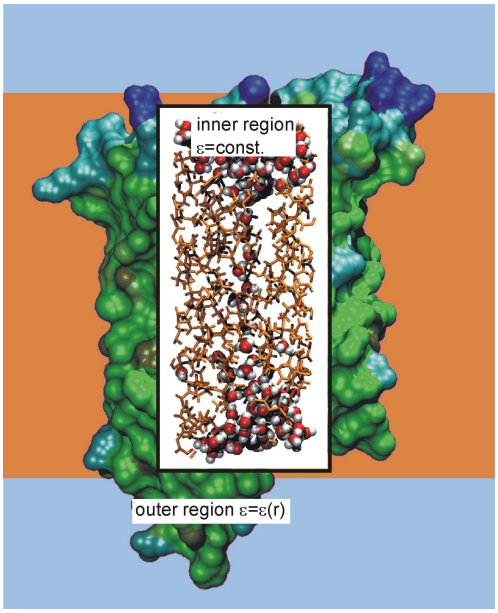


Figure 3.3: Schematic representation of a biomolecular system using the generalized solvent boundary potential. While the atoms in the inner region are treated explicitly, the effect of the atoms in the outer region, the bulk solvent and the membrane if included are taken into account through continuum electrostatics.

instantaneous position of the atomic charges in the inner region and has therefore to be recomputed in every step of the simulation. To circumvent this difficulty, it is useful to express the reaction field $\phi_{rf}^{(i)}$ in terms of the reaction field Green's function $G_{rf}(\boldsymbol{r}, \boldsymbol{r}')$ corresponding to the reaction field potential at $\boldsymbol{r}$ due to the atomic charge at $\boldsymbol{r}'$,

$$\phi_{rf}^{(i)}(\boldsymbol{r}) = \int d\boldsymbol{r}' G(\boldsymbol{r}, \boldsymbol{r}') \rho^{(i)}(\boldsymbol{r}') \tag{3.12}$$

The inner-inner contribution then reads:

$$\Delta W^{ii}(\boldsymbol{r}) = \frac{1}{2} \iint d\boldsymbol{r} d\boldsymbol{r}' \rho^{(i)}(\boldsymbol{r}) G(\boldsymbol{r}, \boldsymbol{r}') \rho^{(i)}(\boldsymbol{r}') \tag{3.13}$$

In order to achieve an efficient calculation, a basis set expansion for the charge distribution in the inner region is used,

$$\rho^{(i)} = \sum_m c_m b_m(\boldsymbol{r}) \tag{3.14}$$

where the expansion coefficients c_m can be calculated using $c_m = \sum_n O_{mn}^{-1} Q_n$. The overlap matrix elements O_{mn} of the basis functions are computed as $O_{mn} = \int d\boldsymbol{r}\, b_n(\boldsymbol{r}) b_m(\boldsymbol{r})$ and the vector of generalized multipole moments Q_n are defined as $Q_n = \sum_{\alpha \in \text{inner}} q_\alpha b_n(\boldsymbol{r}_\alpha)$.

Correspondingly, the inner-inner contribution to the solvation free energy takes the

following form:

$$\Delta W_{elec}^{(ii)}(\boldsymbol{r}) = \frac{1}{2}\sum_{mn} Q_m \left[\sum_{ij} O_{im}^{-1} M_{ij} O_{jn}^{-1}\right] Q_n = \frac{1}{2}\sum_{mn} Q_m M_{mn}^{*} Q_n \tag{3.15}$$

with

$$M_{ij} = \iint d\boldsymbol{r} d\boldsymbol{r}' b_i(\boldsymbol{r}) G(\boldsymbol{r}, \boldsymbol{r}') b_j(\boldsymbol{r}') \tag{3.16}$$

$\mathbf{M}^*$ is termed the reaction field matrix and is the numerical representation of the Green's function for the Poisson-Boltzmann equation in the b_m basis. Similar to $\phi_s^{(o)}$, $\mathbf{M}$ only needs to be computed once via solving the Poisson-Boltzmann equation and does not depend on the instantaneous configuration of the inner region.

The choice of basis functions depends on the geometry of the problem. For membrane proteins, a rectangular boundary is appropriate, while for globular proteins a spherical boundary is often advisable. Accordingly, either Legendre polynomials or spherical harmonics are taken as basis functions.

3.2.2 Generalized solvent boundary potential in the SCC-DFTB/MM framework

As described in a recent work [128], the GSBP approach can be easily extended to a general QM/MM framework. For this, it is assumed that all QM atoms are within the inner region.

The QM/MM contribution to the electrostatic solvation free energy has then the form:

$$\Delta W^{(ii),QM/QM} + \Delta W^{(ii),QM/MM} + \Delta W^{(io),QM/MM} =$$
$$\frac{1}{2}\sum_{mn} Q_m^{QM} M_{mn}^{*} Q_n^{QM} + \frac{1}{2}\sum_{mn} Q_m^{QM} M_{mn}^{*} Q_n^{MM} + \int d\boldsymbol{r} \rho^{QM}(\boldsymbol{r}) \phi_{rf}^{(o)}(\boldsymbol{r}) \tag{3.17}$$

The elements of the vector Q^{MM} differ from the full charge vector due to the fact that modifications required at the QM/MM frontier [212] have to be considered [128]:

$$Q_n^{MM} = \sum_{A \in \text{inner,MM}} \left(q_A + q_A^{shift}\right) b_n(\boldsymbol{r}_A) \tag{3.18}$$

q_A^{shift} contains the charge added onto the MM atom A for the QM/MM treatment as in the SLA, DIV and EXGR schemes [212].

Generally, Q^{QM} and the last term in eq 3.17 depend on the electron density according to the used QM method. Since in SCC-DFTB [138] the QM charge distribution is expressed in terms of Mulliken charges, the calculation of the corresponding GSBP terms is simplified. The generalized multipole moments Q^{QM} are then computed as:

$$Q_m^{QM} = \int d\boldsymbol{r} \rho^{QM}(\boldsymbol{r}) b_m(\boldsymbol{r}) = \int d\boldsymbol{r} \sum_{A \in \text{QM}} \Delta q^A \delta(\boldsymbol{r} - \boldsymbol{R_A}) b_m(\boldsymbol{r}) = \sum_{A \in \text{QM}} \Delta q^A b_m(\boldsymbol{R}_A)$$

Finally, the Hamiltonian matrix elements $H_{\mu\nu}$ of SCC-DFTB are extended to include the GSBP contributions using the following terms:

$$H_{\mu\nu}^{GSBP} = \frac{1}{2} S_{\mu\nu} \sum_{B \in \mathrm{QM}} \left[\Gamma_{CB}(\boldsymbol{R_C}, \boldsymbol{R_B}) + \Gamma_{DB}(\boldsymbol{R_D}, \boldsymbol{R_B}) \right] + \frac{1}{2} S_{\mu\nu} \left[\Omega(\boldsymbol{R_C}) + \Omega(\boldsymbol{R_D}) \right]$$
$$\text{with } \mu \in C, \nu \in D \tag{3.19}$$

where $S_{\mu\nu}$ is the overlap matrix element and the basis functions μ and ν are located on the QM atoms C and D, respectively. The terms $\Gamma_{AB}(\boldsymbol{R_A}, \boldsymbol{R_B})$ and $\Omega(\boldsymbol{R_A})$ are defined as the following:

$$\Gamma_{AB}(\boldsymbol{R_A}, \boldsymbol{R_B}) = \sum_{mn} b_m(\boldsymbol{R_A}) M_{mn}^{*} b_n(\boldsymbol{R_B}) \tag{3.20}$$

$$\Omega(\boldsymbol{R_A}) = \sum_{mn} b_m(\boldsymbol{R_A}) M_{mn}^{*} Q_n^{MM} + \phi_s^{(o)}(\boldsymbol{R_A}) \tag{3.21}$$

They describes the difference between the vacuum and reaction field interaction between QM atoms ($\Gamma_{AB}(\boldsymbol{R_A}, \boldsymbol{R_B})$) and the direct Coulombic and reaction field contribution to the interaction between QM and the outer regions ($\Omega(\boldsymbol{R_A})$).

Chapter 4

Long-Range Proton Transfer in Biomolecular Channels: Developments and a Model System

Proton transfer (PT) processes are an essential component in bioenergetics. Beside active proton transfer and proton transduction for instance in ATPase [232], bacteriorhodopsin [233] or the photosynthetic reaction center [234] and integral membrane proteins such that gramicidin A which rely on the passive proton transport, there are proteins which transport various compounds but prevent PT to maintain the gradient.

Further, many biochemical processes and also technical applications depend on (long-range) PT. Fuel cells for example necessitate well-controlled long-range PTs [235]. In many enzymes, the general acid-base catalysis [236] relies on (local) PT between amino acids and substrates. Long-range PT has been suggested to be responsible for catalytic cooperativity [237] between distant sites.

While the mechanisms important for localized PT in enzyme active sites have been well elucidated [117, 238–243], the investigation of long-range PTs is challenging particularly at a quantitative level. It is generally accepted that hydrogen bonded chains of water molecules, with or without participation of titratable amino acid side chains [244], are involved in long-range PT [86, 245, 246]. But detailed transfer pathway(s) or rate-limiting factors are often difficult to acquire due to the large number of solvent molecules and protein residues involved.

Experimentally, mutagenesis experiments and kinetic measurements for the corresponding mutants provide the means to probe transfer pathways and identify the interaction regulating the transfer kinetics. However, some ambiguity remains in the interpretation since mutations might perturb the solvent or protein structure.

Theoretical investigations can serve as a complementary technique and apart from general electrostatics considerations, the advent of QM/MM techniques has allowed a closer look at the mechanisms involved on an atomistic level while still taking into account the effects of the surroundings. A large number of theo-

retical studies have been conducted for several systems involving long-distance PT [91–93, 106, 107, 109, 247–260] and unraveled useful mechanistic information. However, a quantitative understanding could not be achieved for any of these systems. For aquaporin, for instance, different rate-limiting events were suggested by different groups using diverse simulation protocols [106, 107, 109].

These intricatenesses reflect the technical challenges that have to be solved for generally robust theoretical analysis of long-range PT.

On the one hand, a quantum mechanical treatment of a large number of atoms is necessary for a precise description of the reaction energetics, since a large number of bonds are formed and broken. This becomes especially important if the PT coupled to fairly complex chemical reactions [261]. Previous studies, however, mostly applied empirical models for PT along water wires, which may not be sufficiently accurate or flexible.

Secondly, sufficient conformational sampling of the protein and solvent is increasingly inevitable in order to obtain quantitative estimations of the transfer kinetics, due to large number of shuttling groups explicitly involved in long-range PTs. In this regard, simple minimization approaches have been shown to be problematic [235]. The calculation of minimum energy paths (MEP) might also not be effective, since they provide only a static description of the reaction. A discussion of the caveats of MEP can be found in ref [262]. The potential of mean force (PMF), in contrast, includes the effect of thermal fluctuations, but is also computationally more demanding. Nevertheless, PMF were calculated for the PT in several studies [91–93, 106, 107, 109, 253, 256] but mostly for short-range PT.

Furthermore, the evaluation of PMF along a well-chosen reaction coordinate is an alternative to study rare events such as most PT process. Since PT processes often have significant barriers, they cannot studied with unbiased nanosecond MD simulations [263] which is possible in some systems with fast PT. Other simulation methods such as the transition path sampling technique [264] have been proposed to study the real-time dynamics of rare events in the condensed phase, but their high computational cost limits their applicability.

Aside from these aspects, two more issues emerge and inspired the technical developments introduced later in this chapter.

A rather general problem regards the choice of a suitable reaction coordinate in PMF calculations. Previously either mostly a specific pattern of transfer pathways (e.g., step-wise) were assumed, wherefore all possible mechanisms have to be investigated separately, or the reaction coordinates work best only for transfers across linear water chains. Therefore, a new reaction coordinate has been developed to describe PT along water chains. It is introduced and tested for a model channel in the next two sections.

The second issue concerns the treatment of electrostatics in the highly heterogeneous protein, solvent, and possible lipid environment of complex biomolecular systems. Since long-range PT involves significant charge separation, this aspect becomes particularly important. Although this issue has been addressed in some studies using either Ewald summation [224] or Langevin dipole models [265], many studies, especially those involving QM/MM potentials, use very approximate electrostatics schemes that compromise the reliability of the results. The generalized solvent

boundary approach (GSBP) (see section 3.2.1) is very attractive to solve this problem, since it allows very efficient sampling of the most relevant configuration space for the reaction of interest, without significantly sacrificing important environmental effects. Beside its use as test system for the new reaction coordinate, the model channel is therefore additionally used to study the influence of different dielectric environments on the PT energetics and to illustrate the principles of the recently proposed SCC-DFTB/CHARMM/GSBP simulation protocol [128, 130].

4.1 Reaction Coordinates for Long-Range Proton Transfer

Reaction coordinates are used to describe quantitatively the progress of a given reaction. They are usually specific to the problem investigated but should be able to describe a large variety of different situations. A "well-chosen" reaction coordinate considers the essence of the degree(s) of freedom that mainly regulates the reaction kinetics of the PT. Since both the protein and the solvent reorganize significantly in long-range proton transfer because of the distant charge migration, a proper reaction coordinate can be defined as the energy gap between diabatic states corresponding to localized proton coordinations. A similar approach, the "solvent coordinate" is used for electron transfer in solution. This energy gap coordinate [240, 249, 266] has in fact been applied in combination with EVB potentials [114] to study long-range PT. However, in contrast to the EVB potentials which provide well defined diabatic states, adiabatic QM/MM potentials complicate the definition of an energy gap coordinate. Also, it is less straightforward to apply the energy gap coordinate to simultaneous (concerted) multiple proton transfers.

Therefore, reaction coordinates have been defined based on geometrical properties or on a combination of geometrical and charge properties of the PT reaction. The anti-symmetric stretch coordinate, as pure geometrical coordinate, uses only the distances between the donor, the transferring proton, and the acceptor, while a coordinate based on the location of the center of excess charge (CEC) with respect to the proton donor and acceptor groups combines geometrical and charge properties. Although this implicates a geometrical nature of the coordinate, its dependence on the charge distribution of the translocation pathway ensures that an appropriate reorganization of the environment occurs as its value varies. In fact, the comparison of energy gap and geometrical coordinate yielded similar PMF results [267] for short-range proton transfer reactions. Whether the success applies quantitatively to long-range PT processes remains to be investigated.

Generally, the definition of a geometric reaction coordinate requires to extract those degrees of freedom best describing the reaction. In particular when water molecules are involved in PT, fast interchange occurs [84] between structures which are close to the ideal structures of Eigen- and Zundel ions [268, 269]. The excess proton coordinated to water molecules, that is, the defect, is transfered without any atom actually moving farther than fractions of an Ångstrøm [84, 91]. The challenge for long-range PT is hence to separate these subtle fluctuations responsible for proton transfer from other motion in the immediate protein and solvent environment.

4.1.1 Previous Suggestions

For a short-range PT, in which the location of the transferring proton is well-defined, a commonly used reaction coordinate is the anti-symmetric stretch involving the donor (D), the transferring proton (H), and the acceptor (A),

$$\delta = r^{D,H} - r^{A,H} \tag{4.1}$$

This in principle can be generalized to multiple proton transfers by using a linear combination of anti-symmetric stretch coordinates for all sets of donor, transferring proton, and acceptor atoms. Although such a linear combination was indeed found useful in the study of PT through two and three intervening water molecules in carbonic anhydrase [250], its use becomes cumbersome and less robust for more complicated pathways.

In the study of PT through the water wire in gramicidin A, a reaction coordinate based on the CEC was used by Roux and co-workers [92, 253], which involves the projection of the total dipole moment of the water wire on the z-axis (the axis of the water wire),

$$\mu_z = q_H \sum_{i=1}^{N_H} r_z^{H_i} + q_O \sum_{j=1}^{N_O} r_z^{O_j} \tag{4.2}$$

where N_H and N_O are the total number of hydrogen and oxygen nuclei in the water wire, r_z's are z coordinates, and q_H and q_O are the partial charges of H (+1e) and O (-2e), respectively. The CEC coordinate is hence defined by μ_z/e, where e is the unit charge. For an unprotonated chain of water molecules (e.g., $O_{10}H_{20}$, Fig. 4.1a), the CEC gives the z component of the total dipole moment; in a protonated wire (e.g. $O_{10}H_{21}^+$, Fig. 4.1b-e), on the other hand, it corresponds to the projection of the center of excess charge (defect proton) along the z axis.

In contrast to the anti-symmetric stretch (δ), the CEC coordinate is a *global*, collective coordinate, meaning that it reflects not only the location of the excess proton in a water wire but also the configuration of all the water molecules in the wire. This sensitive dependence on the orientation of individual water molecules makes μ_z/e easily "contaminated" by the fluctuation of the water wire (*vide infra*).

An alternative reaction coordinate suggested by Chakrabarti et al. [106] (denoted by ν in the following), which takes a *local* view at the problem, counts the number (w^{O_i}) of protons coordinated to each oxygen atom O_i in the wire,

$$\nu = \frac{\sum_{i=1}^{N_O} r_z^{O_i} w^{O_i}}{\sum_{O_i} w^{O_i}} \tag{4.3}$$

$$w^{O_i} = \left[\sum_{j=1}^{N_H} f_{sw}(d_{O_i,H_j}) \right] - 2 \tag{4.4}$$

Here and in the following, $d_{A,B}$ denotes the Cartesian distance between atoms A and B. The switching function $f_{sw}(d)$ is given in the following equation with suggested values of r_{sw} =1.4 Å and d_{sw} =0.05 Å [106]:

$$f_{sw}(d) = \frac{1}{1 + exp\left[(d - r_{sw})/d_{sw}\right]} \tag{4.5}$$

Although this coordinate in its optimum yields a more precise description of the location of the excess charge compared to μ_z/e, the limitation is that the information included is purely *local*. As a result, the functional form introduced cannot distinguish between three hydrogen atoms coordinated to an oxygen atom in an oxonium ion and three hydrogen atoms close to an oxygen atom due to a collision of two water molecules. Although this problem can be partially overcome either by choosing a better switching function or by a judicious choice of the parameters in the switching function (eq 4.5), it is difficult to fully eliminate fluctuations of the reactions coordinate due to water collisions. Even though close encounter of water molecules (r <1.4 Å) are not overwhelmingly frequent, the contamination effect of the reaction coordinate is significant as illustrated below.

4.1.2 A New Reaction Coordinate

To take into account the advantages of global *and* local considerations, a new coordinated that unites the formulations discussed above is proposed. This modified CEC (mCEC) coordinate for a linear proton wire (along the z axis), which may include both water and protein groups, is defined as:

$$\xi_z = \sum_{i=1}^{N_H} r_z^{H_i} - \sum_{j=1}^{N_X} w^{X_j} r_z^{X_j} - \sum_{i=1}^{N_H}\sum_{j=1}^{N_X} f_{sw}(d_{X_j,H_i}) \cdot \left(r_z^{H_i} - r_z^{X_j}\right) \tag{4.6}$$

where X_j represents a coordinating atom for protons during the translocation, r_z^A is the projection of the position vector of atom A on the z axis. w^{X_j} is the weight associated with the atom X_j and is defined as the number of protons coordinated to that atom in its reference state. The reference state is the least protonated state of the atom in both reactant and product. For instance, if atom X_k has two protons coordinated in the reactant but only one in the product, then $w^{X_k} = 1$. When the proton transfer is solely through water molecules as shown in Fig. 4.1, the oxygen atoms have a w^X of +2. In the more general case, X can be any proton acceptor with a different weight.

The correction (third term in eq 4.6) consists of the sum over all contributions to the z-component of the bond vector from individual X-H bonds. To establish the definition of a bond in this context, the same switching function $f_{sw}(d)$ as Chakrabarti et al. (eq 4.5) for the distance d_{X_i,H_j} is used. Parameters of $r_{sw} = 1.3$ Å and $d_{sw} = 0.03$ Å were found to give the best results.

4.1.3 Comparison of Different Coordinates for a Linear Proton Wire

To illustrate the difference between the newly proposed ξ_z coordinate and previous suggestions (μ_z/e, ν), their behavior with different situations of protonated and non-protonated water wires (Fig. 4.1) is compared. The coordinates were collected from molecular dynamics simulations of the model channels described below (sec. 4.2).

When there is no excess proton in the water wire, the value of the CEC coordinate (μ_z/e) is significantly different from 0 as expected (Table 4.1) since the dipole moments of the individual water molecules add up. The value of the mCEC coordinate

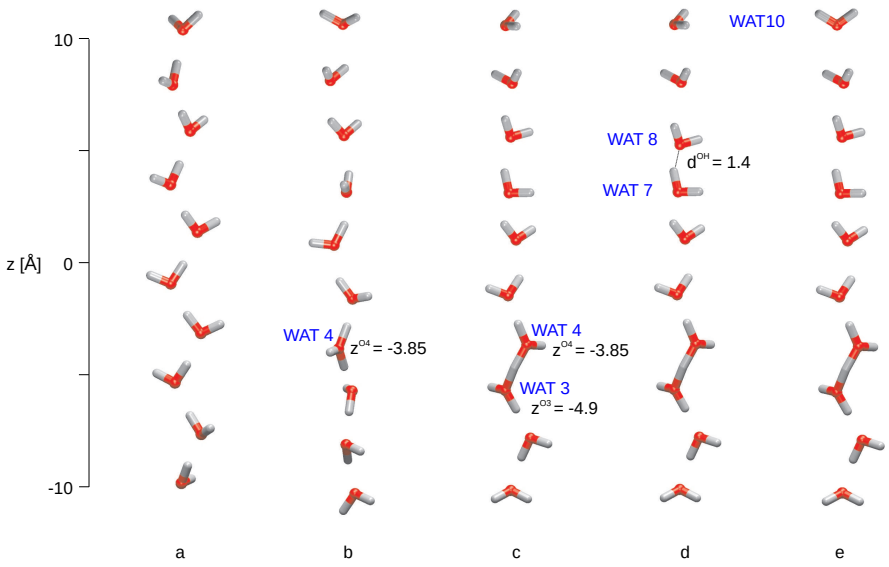


Figure 4.1: Water wire models for illustrating different reaction coordinates that describe long-range proton transfers. a) a water wire without any excess proton; b) a water wire with a H_3O^+-ion located at WAT4; c) a water wire with a Zundel $H_5O_2^+$-ion between WAT3 and WAT4; d) same as c) but with one water molecule (WAT8) displaced to mimic a close collision of water molecules. e) same as c) but with on water molecule (WAT10) rotated by a small angle to illustrate the contamination of μ_z/e. The values of different reaction coordinates discussed here are shown in Table 4.1. The distance cutoff for the bonds drawn was chosen to be 1.3 Å.

(ξ_z), by contrast, is very close to zero; this demonstrates that the correction (third term in eq 4.6) appropriately removes contributions from any components irrelevant to the proton transfer as designed. The reaction coordinate of Chakrabarti et al. [106] (ν) gives unstable results (large numerical value) for unprotonated water wire because the denominator in eq 4.3 approaches zero in this case; this is not a significant shortcoming because no such instability is anticipated for cases with excess proton(s).

For the water wire with a localized hydronium ion (Fig. 4.1b), ξ_z gives the z coordinate of the oxygen atom in WAT4, which is the oxygen that bears the extra proton. Slight fluctuations in the bond lengths are reflected in ν, which is slightly displaced toward WAT3 (Table 4.1). With a Zundel like ion in the wire (Fig. 4.1c), both ν and ξ_z give values between the z coordinate of the oxygen atoms in WAT4 and WAT5. The different distances from the central proton to the two neighboring oxygen atoms (1.3 and 1.4 Å to O4 and O5, respectively) lead to slightly different values of the two reaction coordinates due to different functional forms. With the current set of parameters in f_{sw}, both coordinates locate the excess proton closer to the WAT5 oxygen, which is consistent with the geometry.

The shortcoming of the ν coordinate becomes clear with the case in Fig. 4.1d, in which one water molecule (WAT8) was displaced to simulate a close collision between water molecules. The mCEC coordinate is not affected by this, but the ν co-

Table 4.1: Values for different reaction coordinates (in Å) for the water wires displayed in Fig. 4.1.

	μ_z/e eq 4.2	ν eq 4.3	ξ_z eq 4.6
a	11.4	n/a	0.2
b	-0.5	-3.5	-3.8
c	-0.9	-4.5	-4.2
d	-0.9	-2.9	-4.2
e	-0.3	-2.9	-4.2

ordinate changes significantly: the oxygen in WAT8 appears to have more than two bonded hydrogens, leading to a nonvanishing weight w^{O_i} that modifies the value of ν. While configurations with such close collisions are not frequently sampled, the large impact on the value of ν is devastating to PMF simulations. Moreover, the importance of this problem grows with the system size as the weight of even miniscule collisions increases as a function of its distance to the actual location of the excess proton.

As far as the original CEC coordinate (μ_z/e) is concerned, it does not give a value close to the actual location of the excess proton in any cases in Fig. 4.1. Although this feature by itself does not invalidate using μ_z/e for characterizing long range PTs, a problematic feature of μ_z/e is that it fails to distinguish between degrees of freedom essential for the PT and fluctuations in the environment. To illustrate this point, one water molecule (WAT10) at the end of the single file was manually rotated (Fig. 4.1e). While the coordinates ν and the ξ_z were not affected by this change, the CEC coordinate showed a shift *larger* than that observed in the transition from an oxonium ion (Fig. 4.1b) to a Zundel ion (Fig. 4.1b). This gets even more problematic when larger systems with more water molecules are examined.

The new reaction coordinate shown here for linear chains is further flexible enough for more general PT processes as has been shown for carbonic anhydrase II (CAII) in a joint publication [130].

4.2 A Model Proton Channel

In this test case, PT along a chain of water molecules confined in a model channel is studied. The channel is embedded in different environments described by continuum models to explore the dielectric effect on the PT energetics. The effect of positioning permanent dipoles lining the channel is also briefly studied.

This simple model was inspired by the recent heated discussions regarding why some membrane water channels conduct protons whereas others do not (see, e.g., a nontechnical summary in ref [270]). For instance, PT through gramicidin is rather facile. In fact, this is the functional mechanism of this antibiotic peptide. By contrast, aquaporins, efficiently conduct water molecules [8, 88] yet exclude protons and other charged ions from passing through. Different factors have been suggested to

determine proton conductance, which include structure of water wires inside the protein [88, 89], electrostatic interaction between the protein and the proton [105, 108], the dielectric barrier [109], or combinations thereof [106]. A detailed overview over the topic will be given in chapter 5.

Construction of the Model and Computational Details

The overall approach is based on MD simulations in a combined QM/MM framework [136] using the approximate density functional method SCC-DFTB [138] for the QM region and the CHARMM27 force field [137] for the remainder of the MM region. To incorporate the effects of solvation and a membrane environment, the generalized solvent boundary potential (GSBP) approach, described in section 3.2.1, was used in the SCC-DFTB/MM framework [130].

The model is similar to the one used in a previous study [253] and was constructed using the following procedure. A linear water chain consisting of 10 water molecules and one excess proton was built, in which the distance from the first to the last oxygen is 22.0 Å; the choice of the number of water molecules and the length of the chain was motivated by the structures of aquaporin and gramicidin A, in which the single-file part of the transmembrane water chain typically consists of 9-10 water molecules. The model "channel" that holds the water wire was introduced by applying a cylindrical potential to each of the oxygen atoms in the water wire:

$$E_{cyl} = \frac{1}{2} k_{cyl} (r - r_0)^2 \Theta(r - r_0)$$

with k_{cyl} =100 kcal mol^{-1} Å^{-2}. The quantity r is the distance of the oxygen atom from the axis of the channel, and r_0 is 0.5 Å; Θ is the Heaviside step function. To ensure appropriate solvation of the excess proton at the ends of the water chain, fifteen water molecules were added to each end of the chain. The coordinates of these water molecules were taken from an equilibrated box of water molecules at 300 K. The "bulk" water molecules were subject to a set of quartic constraints to maintain a cubic shape of 3×3×2.5 Å.

This "dumb-bell" model system was then embedded into different dielectric environments including vacuum, pure solution, and membrane systems. Important geometrical parameters are illustrated in Fig. 4.2 and their values are summarized in Table 4.2 for different environments. In the membrane set-ups, the grids used in GSBP and midpoint of the membrane was centered on the model channel. The grid dimensions were 92×92×185, and a coarse grid spacing of 1.6 Å and a finer spacing of 0.4 Å were used in focusing Poisson-Boltzmann (PB) calculations required for the various quantities used in the GSBP computations. The dielectric constants for the membrane and bulk water were set to be 2.0 and 80.0, respectively; although the PB module in CHARMM allows the use of different dielectric constants for the lipid polar heads and nonpolar tails, this was not pursued here. The salt concentration was chosen to be 150 mM. Although Im et al. [271] showed that the artifacts encountered with trilinear interpolation should have a minor effect on molecular dynamics simulations, the slightly more involved cardinal B-spline method for interpolating between the gridpoints were used. In the GSBP calculations for the inner electro-

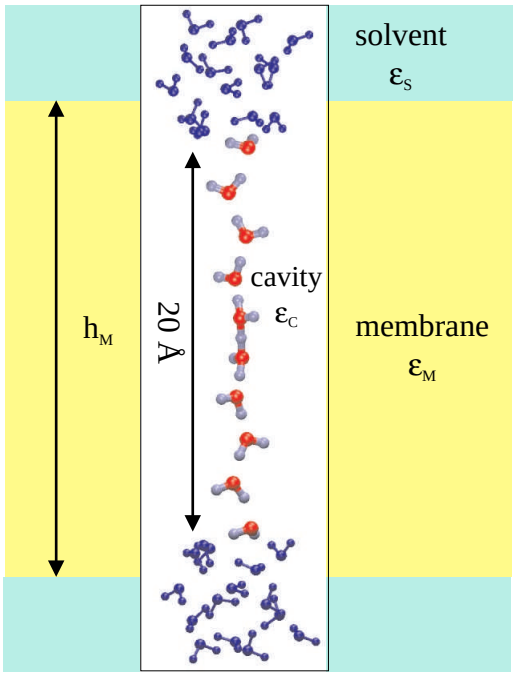


Figure 4.2: Geometric and dielectric parameters for the model channel. See Table 4.2 for specific numerical values for the dielectric constants (ϵ_C, ϵ_M, ϵ_S) and the membrane thickness h_M.

Table 4.2: Geometric and dielectric parameters for the model channel shown in Fig. 4.2: dielectric coefficients for cavity ϵ_C, membrane ϵ_M and solvent ϵ_S and membrane thickness h_M.

	ϵ_C	ϵ_M	ϵ_S	h_M [Å]
vacuum	1	1	1	n/a
water	1	80	80	n/a
membrane 20	1	2	80	20
membrane 30	1	2	80	30

statics, Legendre polynomials up to order 10 were used as basis functions, which was found sufficient in previous studies for systems of similar size [135, 272].

To study the effect of polar residues on the PT energetics, dipoles were added in the lumen of the model channel embedded in the 30 Å model membrane. Specifically, four dipoles with rather modest dipole moments composed of two opposite charges ($|q|$=0.1e, d=1.5Å) were circularly arranged around the channel at a radius of 3.3 Å. Three of these annuluses were positioned at z=-1.4,1.2, 3.8 Å (Fig. 4.4), which were motivated by the carbonyl groups in gramicidin A; two extreme orientations of the dipoles were considered.

The chain of water molecules and the excess proton were described with the SCC-DFTB method, while the rest explicit water molecules were described with the TIP3P model. Additionally, calculation with all water molecules described by SCC-DFTB

were conducted in order to evaluate the validity of the QM/MM approach. PMF calculations were carried out using the standard umbrella sampling technique [273] with ξ_z as the reaction coordinate since the water wire here is highly linear. A total of ten windows was used, and each window contained 40 ps of equilibration and 100 ps of production calculations.

Results and Discussion

In vacuum, the proton transfer in the dumb bell model has a modest barrier of 6.3 kcal/mol (Table 4.3, Fig. 4.3); there is a barrier because the excess proton is preferentially stabilized by the explicit solvent molecules at the ends of the water wire. The barrier increases slightly to 8.1 kcal/mol, when the model is immersed in bulk water represented using a dielectric medium of $\epsilon = 80$. With the dielectric model for the membrane, the effect of preferential solvation at the ends of the water wire becomes more significant, which causes the PT barrier to increase even further; with the membrane thickness (or low-dielectric part of the membrane) of 20 and 30 Å, respectively, the barrier is 10.3 and 9.8 kcal/mol, respectively.

The results shown here are in qualitative agreement with electrostatic calculations of Burykin et al. [109], who found a barrier of about 15 kcal/mol for transferring a proton through a water wire in a cylindrical pore of radius 4 Å through a 30 Å thick membrane. However, the barrier observed here includes only part of the dielectric barrier due to the membrane. Since the initial position of the excess proton is close to the low-dielectric region in the current model, the charge has already been partially desolvated. To include the entire dielectric barrier, the free energy necessary for transporting the proton from the bulk to the minimum in the PMF has to be computed [274].

The calculation of the PMF using SCC-DFTB for all explicit water molecules yielded barriers heights which are slightly increased compared to those obtained with the QM/MM approach (Table 4.3). The differences range from 0.2 kcal/mol for the channel embedded in vacuum to 1.2 kcal/mol using a membrane with the thickness of 30 Å.

Table 4.3: Barrier Heights $W^{\ddagger}$ (in kcal/mol) for the proton transport in the model channel obtained using different dielectric environments.

	QM/MM[a]	QM[b]
vacuum	6.3	6.5
water	8.1	8.7
membrane 20	10.3	10.9
membrane 30	9.8	11.0

[a] The QM region consists of the chain of water molecules and the proton, remaining water molecules are treated molecular mechanically.

[b] The QM region consists of all explicit water molecules and the proton.

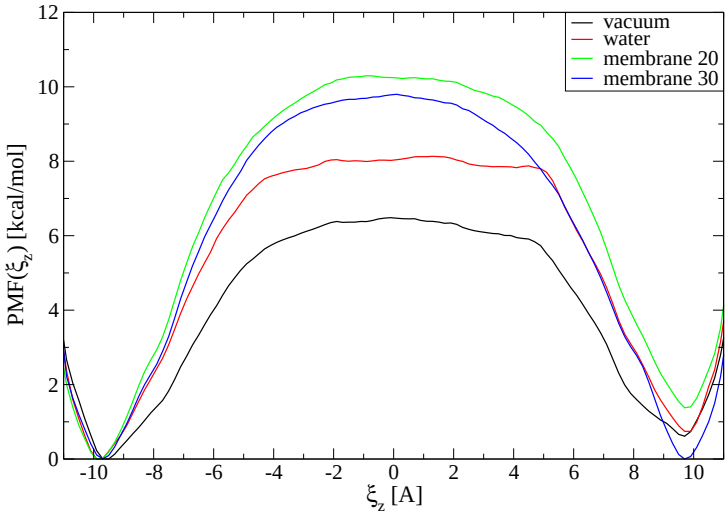


Figure 4.3: Potential of mean force for proton transfer through a purely nonpolar model channel embedded in different dielectric environments.

Generally, the description of the solvation of a proton at the end of the wire is improved using a full QM approach instead of a QM/MM scheme, since collective effects play an essential role for the description of protonated water clusters. A description of the water molecules in the cap using SCC-DFTB therefore results in an increase of the proton affinity and in turn in an increase of the barrier for the PT through the channel. The energy difference of the barrier heights between a full QM and a QM/MM calculation thus indicates the influence of collective effects and electronic polarization. The slight increases of the barriers obtained here (Table 4.3) therefore suggest that the solvation of the proton at the ends of the water wire is sufficiently accurately described in the QM/MM simulations.

The presence of the dipoles along the channel was found to have a major impact on the PT energetics, which was qualitatively expected. With the positive charges of the dipoles facing the center of the channel, the PT barrier is 23 kcal/mol (Fig. 4.4). In contrast, the opposite polarity lowers the PT barrier dramatically from $\sim$ 10 kcal/mol for the nonpolar channel to nearly vanishing. In fact, the most stable configuration corresponds to the excess proton localized in the center of the channel close to one sets of dipoles ($z \sim 1.2$ Å).

In short, study of this simple model showed that the GSBP approach is able to capture the qualitative effect of different environments on the PT energetics along a chain of water molecules. The transition from a high dielectric medium to a low dielectric medium modeled in the slab model described above imposes a notable effect on the PT energetics, on the order of a few kcal/mol. However, the effect of the polar groups mimicked by a set of dipoles of rather small magnitude has a more significant effect. Hence, the results of this simple model support the claim of Warshel and co-workers that nonpolar membrane channels per se do not conduct protons, *unless* the charged species can be stabilized within the channel as described recently [109].

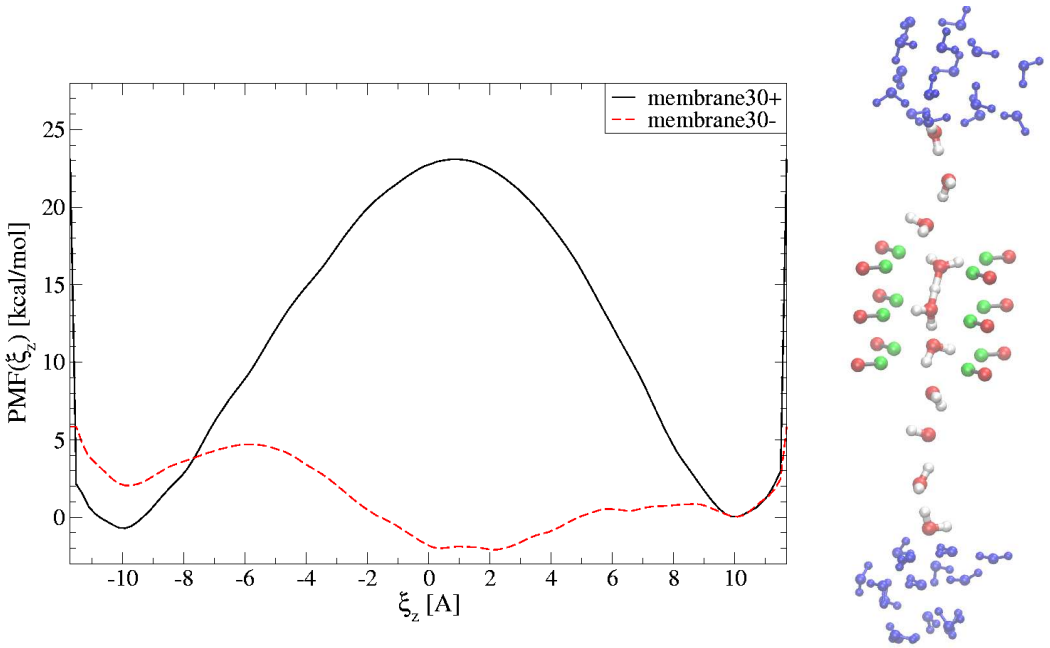


Figure 4.4: Potential of mean force for proton transfer through the model channel with dipoles arranged in the channel. The solid (dashed) line refers to a setup where the positive (negative) charges of the dipoles point to the center of the channel.

4.3 Concluding Remarks

For characterizing the progress of long-range PT, a new reaction coordinate (mCEC) was introduced and successfully applied. The new reaction coordinate unites several advantages: the problems encountered by earlier suggestions which were also analyzed here are eliminated, it works without assuming a priori the pathway, and in contrast to earlier suggestions a generalization to three-dimensional transfer pathways is easily possible as shown in a joint publication [130].

The applicability of the new coordinate was verified by calculating the PT reaction in a model channel and by other authors in systems, such as carbonic anhydrase [130]. The results suggest that it is perfectly suited as reaction coordinate for computing meaningful potential of mean force for long-range PT.

The results for the influence of electrostatics in long-range PT through the model channel demonstrated the applicability of the QM/MM/GSBP simulation protocol for the simulation of long-range PT in heterogeneous systems. In particular, the comparison of the QM/MM with full QM results also showed that the usage of a QM/MM approach to simulate PT trough channels in a complex dielectric environment is justified.

Furthermore, the model channel helped to understand two aspects of membrane channels: the dielectric barrier and the tuning of conduction properties using polar groups lining the channel interior.

Chapter 5

Proton Exclusion in the Glycerol Facilitator GlpF

The glycerol facilitator GlpF is the second aquaporin in the cell membrane of *Escherichia coli* beside the classical aquaporin AqpZ. GlpF is an aquaglycoporin that highly selectively conducts glycerol, water, and other small, uncharged organic molecules [8, 69, 80, 82, 275, 276] but is impermeable to protons and other ions [277]. Glycerol is an important metabolite and component of $\sim$2/3 of all phospholipids. Inside the cell, it is phosphorylated by glycerol kinase thus maintaining the concentration gradient that drives inward glycerol transport.

GlpF had been functionally characterized [80, 278] long before aquaporin-1 (AQP1) was discovered [66–69]. In fact, the existence of a glycerol conducting channel was deduced about one hundred years ago by Alfred Fischer [279]. After the discovery of AQP1, the close relationship of the sequences was recognized and GlpF subsequently classified as aquaglycoporin.

The importance of GlpF as model system is twofold: It provides the possibility to investigate the mechanisms of the water/proton selectivity in aquaporins. On the other hand, the availability of a classical aquaporin (AqpZ) and an aquaglycoporin (GlpF) in the same bacteria provides the possibility to investigate the water/glycerol selectivity unaffected by species or tissue differentiation.

In the next section, the structure of GlpF and the mechanism and selectivity of the glycerol conductance are shortly reviewed, followed by an outline of the current state of knowledge on the mechanisms of the proton blockage (sec. 5.2). In section 5.3, the mechanism of proton exclusion in GlpF is investigated. Concluding the results are summarized in section 5.4.

5.1 Structure and Glycerol Conductance of GlpF

By now, the atomic structure of seven members of the aquaporin family is known - AqpZ [96] and GlpF [82, 88] from *Escherichia coli*, the mammalian AQP1 [97, 100, 101, 280], AQP0 [98, 99] and AQP4 [281], the plant aquaporin SoPIP2 [282], and the archaeal aqua-

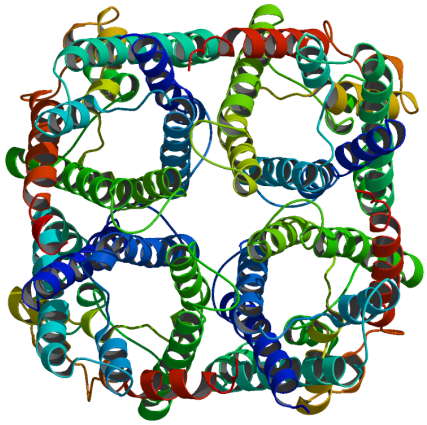


Figure 5.1: Structure of the GlpF tetramer in the membrane.

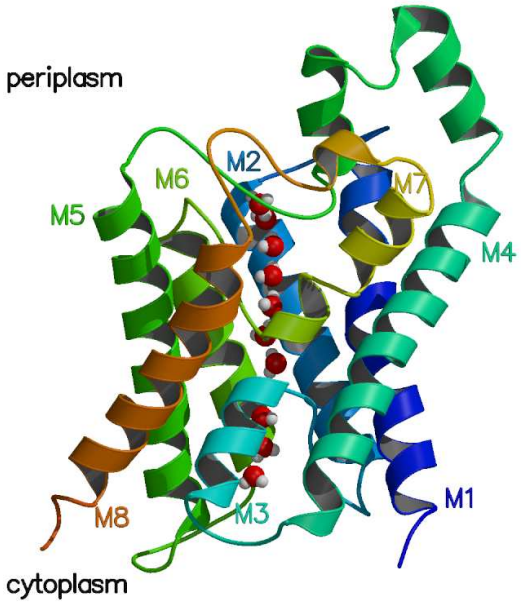


Figure 5.2: Structure of a GlpF monomer viewed down the quasi twofold symmetry axis. The line of water molecules through the pore is shown in CPK representation.

porin AqpM [283]. The atomic structures and information from the sequences of a large number of aquaporins [284] revealed several highly conserved core segments forming a protein structure which is common to all aquaporins. In the following, the structure of GlpF is thus also exemplarily discussed for all aquaporins.

GlpF is embedded into the membrane as a homotetramer with a four-fold symmetry axis normal to the membrane plane [82, 285] (Fig. 5.1). Each monomer contains a pore/channel surrounded by six bilayer-spanning α-helices (M1, M2, M4, M5, M6 and M8) and two half-spanning helices (M3 and M7). Together the helices are arranged in a right-handed twisted bundle (Fig. 5.2).

The sequence of GlpF shows an internal repeat - the N-terminal segment has $\sim$20% conservation with the C-terminal segment [286]. This internal similarity was suggested to arose from a tandem gene duplication [287] and is also reflected in the three-dimensional structure. The two segments (helices M1 to M4 and M5 to M8) are related by a quasi twofold axis that lies in the mid-membrane plane [82, 285] (Fig. 5.2). Such a structural two-fold relation was also found, for instance, for CIC chloride channels [288] or the lactose permease LacY [289].

Each segment contains three bilayer-spanning α-helices and a functional loop, connecting the second and the third transmembrane element (see Fig. 5.3 for the N-terminal segment). The N-terminal segment starts at the cytoplasmic side with two membrane-spanning α-helices, M1 and M2, followed by a stretch of seven amino acids that folds into the membrane as an extended polypeptide chain and orients three successive backbone carbonyl groups into the channel (Fig. 5.3). This chain

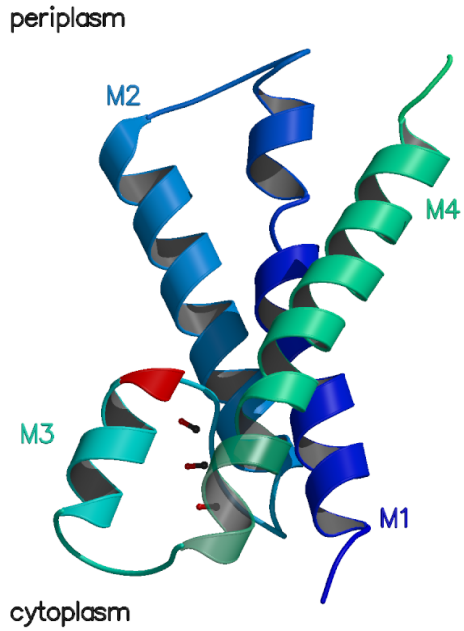


Figure 5.3: Structure of the N-terminal segment of GlpF. The carbonyl groups of extended polypeptide chain lining the pore are highlighted in ball-and-stick representation. The position of the NPA motif is colored in red.

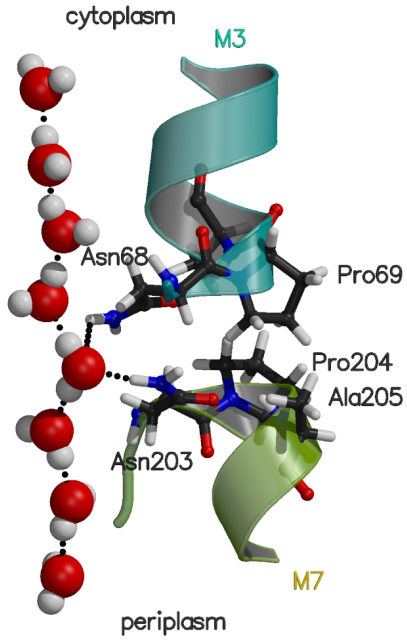


Figure 5.4: Structural details of the NPA motifs region at the interface of the half-spanning helices M3 and M7. Shown are the two NPA motifs and the hydrogen-bonded water network with its bipolar orientation starting from the central water molecule hydrogen-bonded to the NPA motifs.

returns beginning with a highly conserved -Asn-Pro-Ala- (NPA) sequence near the center of the membrane as one of the half-spanning helices, M3, back to the cytoplasmic side. The third transmembrane helix, M4, is located between M1 and M3.

The C-terminal segment shows a similar transmembrane fold beginning from the periplasmic side. The transmembrane helices M5 and M6 are followed by the second extended polypeptide chain which enters the membrane from the periplasm, also orients three backbone carbonyl groups into the lumen and initiates the second half-spanning helix M7 at the center of the membrane beginning with the second NPA motif.

Because of this special topology, the two functional loops touch midways in the membrane with the N-terminal ends of their two half-spanning helices bringing the two conserved NPA motifs in close contact (Fig. 5.4). The proline rings are in van-der-Waals contact to each other and placed between the proline and alanine side chain of the other segment. The asparagine side chain of the NPA motifs are oriented into the channel and act as hydrogen bond donors (Fig. 5.4). The carbonyl groups of the extended polypeptide chains form two ladders of hydrogen bond acceptors one on each side of the NPA motifs (Gly199, Phe200, Ala201 on the cytoplasmic side; Gly64, Phe65, Ala66 on the periplasmic side) (Fig. 5.5).

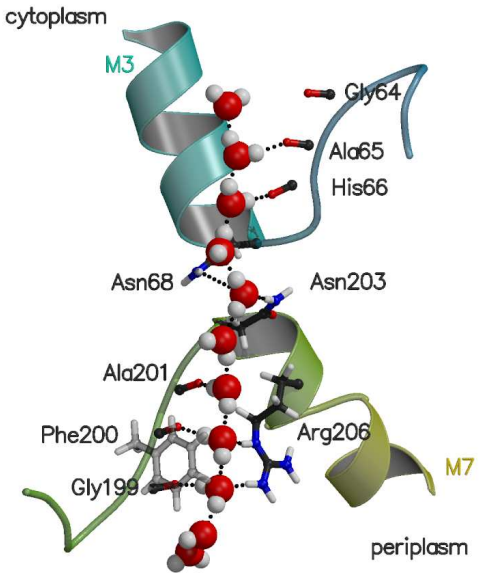


Figure 5.5: Structural details of the conduction pathway and the water chain through the pore of GlpF. Depicted are the water chain, the half-membrane spanning repeats with helical (M3 and M7) and nonhelic parts, the carbonyl groups lining the pore and the side chains Arg206 and Phe200 of the selectivity filter and Asn68 and Asn205 of the NPA motifs.

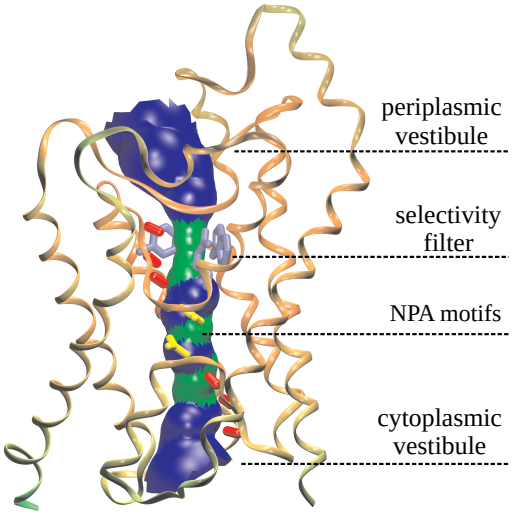


Figure 5.6: The shape of the GlpF pore. The lining of the pore is shown as solid surface. The selectivity filter (blue), the carbonyl groups lining the pore (red) and asparagine residues of the NPA motifs (yellow) are shown in ball-and-stick representation.

Structure of the Channel Pathway

The channel pathway consists of three elements: a periplasmic vestibule, an extended narrow pore, and a cytoplasmic vestibule [82, 97, 285, 290]. The periplasmic vestibule tapers down to its constriction of ∼3.8 Å by 3.4 Å at a motif called selectivity filter (SF) and continues as a ∼28 Å-long narrow pore (radius < 3.5 Å) that following widens out toward the cytoplasmic side (Fig. 5.6)

The SF is located about 8 Å above the quasi twofold axis to the periplasmic side and consists of two aromatic side chains (Trp48, Phe200), the guanidinium side chain of Arg206, and two backbone carbonyl groups (Gly199, Phe200) (Fig. 5.5). The cross section is precisely large enough to allow a single CH-OH group to pass. Therefore, glycerol molecules (and water molecules) can only pass in a single file.

Because of its composition, the SF is strongly amphipathic. The two aromatic side chains form a hydrophobic corner on two sides of the pore, while the guanidinium side chain of Arg206 constitutes the third side and provides hydrogen bond donors. On the forth side, the two backbone carbonyl groups act as hydrogen bond acceptor sites. This amphipathic property is continued in the narrow pore - the line of accessible carbonyl groups and the highly constrained asparagine side chain of the NPA motifs provide hydrogen bond acceptor/donor sites along the path while the

opposing surface of the pore is mainly hydrophobic.

The amphipathic nature also determines how the glycerol molecules are conducted. At the SF, for instance, the alkyl backbone of the glycerol is tightly packed against the hydrophobic corner, leaving no space for substitutions at the C-H hydrogen positions [82, 285]. The hydroxyl groups of the glycerol act as hydrogen bond acceptors and donors from the guanidinium side chain and the two backbone carbonyl groups, respectively.

Selectivity

As shortly outlined above, GlpF allows the permeation of water, glycerol but also of other linear polyalcohols [80, 82] (alditols). Cyclic compounds are impeded by their size since the channel diameter is too small [80, 82].

The permeation of different alditols through GlpF is stereoselective. For example, ribitol, a five-carbon alditol, shows a similar rate of transport as glycerol, while D-arabitol, a stereoisomer of ribitol, exhibits a ten-fold reduction in transport [82] relative to ribitol.

The explanation for this selectivity was suggested to be based on the structure and the amphipathic nature of the channel [82, 285]. Due to the narrow dimensions of the pore, the carbon backbone of any alditol must be lined up along the channel axis. At the SF the carbon backbone of two successive CH-OH groups is thereby in close contact with the hydrophobic corner, while the hydroxyl groups are hydrogen bonded to Arg206 and the two carbonyl groups. For any adjacent CH-OH group to these two groups, there are two tetrahedrally different positions with different environments to place the carbon [82].

Therefore, alditols which can align along the channel are conducted. However, the conduction is less efficient if adjacent carbon atoms are placed in the position that mostly clashes with the channel wall. Since in D-alditol the carbon atom is in the position that mostly clashes with the channel wall, the conduction is hindered. Ribitol, in contrast, can adapt a conformation similar to glycerol.

Both, aquaglycoporins and classical aquaporins such as AqpZ conduct water but classical aquaporins do not conduct glycerol. The differences in glycerol permeability was suggested [82, 97] and recently theoretically demonstrated [291] to be caused by the different channel diameter which is larger in GlpF than in AQP1 or AqpZ. This is especially true for the SF whose diameter is significant smaller in classical aquaporins and whose nature is more hydrophilic. The replacement of the hydrophobic corner of the SF in GlpF removes the favorable interaction with the alkyl chain of glycerol and thus together with the reduced diameter impedes the glycerol conduction.

In absence of glycerol, the conduction of water in GlpF is fivefold less efficient than in the classical aquaporin AqpZ [81]. As cause for the reduced rate in GlpF, the hydrophobic nature of the SF was suspected [82, 290]. However, experimental evidence is sparse [81]. For high concentrations of glycerol, on the other hand, crystal structures exhibit three bonded glycerol molecules plus several water molecules between them, suggesting that both species are co-transported [82, 292]. In fact, theoretical re-

sults suggested that water is required for glycerol transport, as it is part of the transport mechanism [292].

5.2 Proton Blockage in Aquaporins - Current State of Knowledge

The attempts to address the water/proton selectivity in aquaporins can be divided into two different categories. The underlying conceptual views rest either upon a geometrical or an electrostatic origin of the selectivity. The studies supporting the first view implicitly assume that an one-dimensional version of the Grotthuss mechanism [245] controls the migration of protons in a column of water molecules forming a water wire in a channel. The primary control of the proton transfer is then provided by a disruption of the specific arrangement of the water molecules, which is prerequisite for proton exchange [87, 245]. In contrast, studies supporting the second view propose that electrostatic interactions dominate the proton exclusion mechanism. The state of knowledge is reviewed in refs [270, 293] and most recently in ref [294].

Structure Based Explanations

Based on the analysis of atomic resolution structures obtained from X-ray diffraction and electron crystallographic data, two different explanations were proposed. Ren et al. [100] suggested based on their structure without resolved water molecules a spacial disruption of the continuous water chain through the channel. Indeed, Sui et al. [295] could not find a suitable chain of hydrogen bonded water molecules in their high-resolution X-ray structure of AQP1. In contrast, Murata et al. [101] proposed an orientational disruption of the continuous chain of water-water hydrogen bonds in form of a hydrogen-bond isolation mechanism. According to their suggestion, the migration of protons is thereby blocked at the NPA motifs, because the hydrogen atoms of the water molecule which is hydrogen bonded to the NPA motifs are oriented perpendicular to the channel axis thus preventing them from forming hydrogen bonds with adjacent water molecules.

Both mechanisms, the spatial and a revised orientational disruption mechanism, were supported by various theoretical studies. Molecular dynamics [89, 90] and Monte Carlo simulations [104] of the water permeation through AQP1 and GlpF yielded a disruption of the water chain in the channel center or the region of the selectivity filter.

In contrast, other studies attributed the proton exclusion to the noncontinuous hydrogen bonding chain through the channel [88, 105]. Based on MD studies of GlpF [88, 105, 296] and AQP1 [89], a revised view of the original proposal [101] suggest a bifurcation of the hydrogen bonding network around the NPA motifs. In this picture, the Asn residue of the NPA motifs act as hydrogen bond donors to a single water molecule restraining its dipole moment perpendicular to the channel axis and preventing it from acting as hydrogen bond acceptor for other water molecules. As a result of the water orientation at the NPA motifs, the water molecules in the two

halves of the channel are oriented such that successive O-H bonds are oriented toward the respective exit of the channel (see Fig. 5.5). This bipolar arrangement contrasts with the monopolar orientation in a water wire necessary to conduct protons by means of the Grotthuss mechanism. In channels observed to conduct protons, e.g. gramicidin, MD studies have revealed a monopolar orientation of the water chain [91, 92].

The bipolar water orientation was demonstrated to be stabilized by the electrostatic interaction of the water molecules with the channel environment [88, 105–107], which thus interferes with the formation of a monopolar water wire. In particular, the helix macrodipoles of the half-spanning helices M3 and M7 and the side chains of the Asn residues of the NPA motifs were shown to be of crucial importance. The carbonyl groups lining the pore on the other hand were shown to be of minor importance [105] but stabilize the position of the water molecules in the channel due to the hydrogen bonding to the water molecules (Fig. 5.5).

However, the studies of the water permeation through AQP1 [89, 90] and GlpF [88, 105] did not incorporate an excess proton in the channel and therefore could only provide indirect evidence regarding the mechanism of water/proton selectivity.

The spacial disruption mechanism, on the one hand, is unlikely to be the general mechanism of proton exclusion since counter-examples showing a continuous water chain throughout the simulation could be found [88, 105]. Other studies found a bipolar water orientation [88, 89, 105–107]. However, thermal fluctuations were shown to readily allow the formation and disruption of hydrogen bonding patterns in the channel of GlpF and AQP1 [104] challenging the view of the bipolar orientation as main source of proton exclusion.

Explanations Based on Explicit Description of Proton Transfer

Recent simulations of GlpF and AQP1 involved an explicit treatment of excess protons [104–112] and used a wide variety of methods to investigate the mechanism of the proton exclusion. All studies consistently predict that the highest point of main barrier against proton permeation is located in the center of the channel and in contrast to the structure based explanations of electrostatic origin. Yet concerning barrier heights and the factors dominating the electrostatic barrier the studies came to different results. The predicted barrier heights range from 28 kcal/mol [112], 26 kcal/mol [111], 15 kcal/mol [109], 12-15 kcal/mol [104], and 6-7 kcal/mol [108] for AQP1 and 19 kcal/mol [110] and 11 kcal/mol [106] for GlpF reflecting the diversity of the applied methods. Regarding the nature of the electrostatic barrier, two quite different conclusions persist: a domination of the barrier due to the electrostatic field generated by the protein environment or due to electrostatic desolvation effects.

Explanations based on the electrostatic interaction with the protein environment

The explanation of the proton exclusion due to electrostatic interaction with the protein environment is mainly supported by studies from de Groot et al. [108] and Chakrabarti et al. [106, 107]. In both studies, free energy profiles for the migration of a proton through AQP1 [108] and GlpF [106, 107] were compared with the corresponding electrostatic potential. Because of the qualitative correspondence of the profiles,

in particular the profiles correspondingly peak at the NPA region, electrostatic interactions with the channel environment were concluded to be the dominant factor in the proton exclusion mechanism in both studies. As major contributing factors to the barrier, the macrodipoles of the helices M3 and M7 and the partial charges of the NPA motifs were identified. The influence of the low-dielectric membrane environment, however, was suggested to be of minor importance [106, 107].

These results indicated that electrostatic interactions with the protein matrix instead of the bipolar orientation of the water molecules itself oppose the proton transfer. The stringent control of the water orientation is rather determined by the electrostatic environment [106, 107] as already outlined above. Interestingly, a second smaller barrier [108, 110] or shoulder [106, 107] were found at the selectivity filter region and explained with the spacial disruption mechanism [108].

Based solely on classical electrostatic calculations, Jensen et al. [105] also expected in their investigation of the electrostatic tuning of permeation and selectivity in GlpF the main barrier located at the NPA motifs. However, the proton exclusion was nevertheless associated with the bipolar water orientation (see Structure Based Approaches).

A completely different point was put forward in a very recent study by Chen et al. [112], who suggested that the selectivity filter is also essential for the proton/water selectivity of aquaporins. The simulations are based on a recent experimental work by Beitz et al. [102], in which point mutations were introduced into the selectivity filter of AQP1. Interestingly, proton transfer through AQP1 could be demonstrated for the R195V single mutant and the R195V/H180A double mutant. The calculated free energy profiles [112] for the mutants showed for the mutants R195V and R195V/H180A a drastic drop of the secondary barrier at the SF region similar to previous calculations [107] investigating the influence of the charge of the Arg. The main barrier is also significantly reduced, from 28 kcal/mol to 13 and 17 kcal/mol for the single and double mutant. Additionally, the authors investigated the H180A single mutant, for which also a substantial reduction of the main and secondary barrier was demonstrated, which was suggested to result from a reduction of the dehydration penalty.

Explanations based on the desolvation effects

As alternative to the explanation that the barrier is directly attributed to the electrostatic field generated by the protein matrix, the electrostatic barrier for transferring a charge through a low dielectric region was proposed as main factor determining the selectivity in aquaporins [109, 111].

This explanation is mainly supported by investigations of Burykin et al. [109, 111]. A comparison of free energy profiles for the proton migration considering only contributions of the solvation energy * and in the second case additionally of the charged residues led to the suggestion that the main factor controlling the proton migration is the contribution of the solvation energy. Based on an analysis of the corresponding EVB profile, which additionally incorporates the chemical effect of the proton transfer between neighboring water molecules, a control of the PT due to the de-

*This includes contributions of the permanent dipoles of the protein and the helix dipoles.

tailed orientation of the donor and acceptor was excluded [109]. An analysis of the factors contributing to the barrier later demonstrated [111] that in contrast to the results of de Groot et al. [108] and Chakrabarti et al. [106, 107], the contribution of NPA motifs to the barrier is quite small.

Interestingly, the authors further demonstrated [109] that the free energy profiles for the movement of a proton through the AQP1 pore and a hydrophobic model channel are quite similar. This result was interpreted to further support the explanation that the dielectric properties of the protein and its environment and therefore the desolvation penalty of moving the proton charge from bulk solution to water molecules in the channel where the charge is less solvated determine the electrostatic barrier.

The importance of dehydration penalty of the proton was also suggested by Miloshevsky et al. [104]. But additionally, they concluded that the residues of the protein also significantly contribute to the proton blockage.

So the theoretical studies came to different conclusions about the nature of the selectivity in aquaporins. The explanations based on structural considerations have been demonstrated to be incompatible with the simulations which explicitly incorporate the proton transfer reaction. However, the current consensus explanation of an electrostatic origin of the selectivity is disputed regarding the nature of the electrostatic barrier – either a barrier due to electrostatic interactions with the protein matrix, mainly NPA motif, helix dipoles or selectivity filter, or due to desolvation effects.

But,in this regard, it must be considered that the discrepancies could well originate in the wide variety of used computational methods and simulation schemes. The employed computational methods vary from classical electrostatic calculations [105, 106, 108] and the nondeterministic stochastic hopping method Q-HOP [108] over empirical [104, 106, 107, 109, 111] and semiempirical methods [110, 112] to an ab initio description of the water wire [297]. While the latter in contrast to the other methods took into account the electronic structure of the water wire, the computational demands of these Car-Parrinello MD simulations, however, prohibited the calculation of free energy profiles for the proton migration [297]. Moreover, the use of finite size models [106, 107, 110] or the neglect of the membrane environment [106, 107, 110, 112] are potential sources of discrepancies. A rigid treatment of the channel [104, 105] or the inherent neglect of thermal fluctuations in Poisson-Boltzmann calculations [106] further increase the ambiguity.

5.3 The Mechanism of Proton Exclusion in GlpF

In the following, the results of calculations on the proton exclusion in GlpF are presented. After a short overview over the computational details (sec. 5.3.1), simulations on the water distribution in GlpF using the GSBP-QM/MM simulation protocol are presented in section 5.3.2. Following in section 5.3.3, the mechanisms of the proton exclusion in GlpF are explored based on free energy calculations of the proton transfer reaction through GlpF.

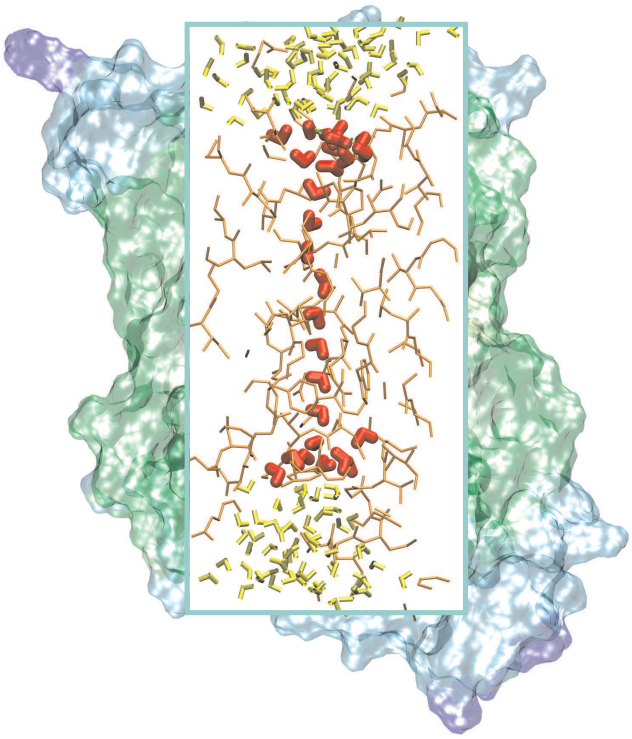


Figure 5.7: Partitioning of the aquaporin system for QM/MM-GSBP simulations. The inner rectangular box shows the atoms located in the inner cavity in the GSBP set-up. The atoms treated using SCC-DFTB are highlighted in red. Atoms in the outer region were fixed during the simulation and their effect were taken into account through continuum electrostatics (see text).

5.3.1 Computational Details

The QM/MM-GSBP simulations were initiated from a snapshot from the classical MD simulation of a GlpF tetramer embedded in solvent and a membrane of Tajkhorshid et al. [88]. Only one monomer was selected for simulation and the rest proteins and explicit membrane were discarded. As shown schematically in Fig. 5.7, the GSBP set-up partitioned the system into inner and outer rectangular cavities that contain 2473 and 1828 atoms, respectively. The size of the inner region is $22\times25\times54$ Å^3, which is described by a set of grids with dimensions of $139\times161\times175$ in the PB calculations for the various GSBP components. A coarse grid spacing of 1.6 Å and a finer spacing of 0.4 Å were used in focusing Poisson-Boltzmann (PB) calculations. The coordinate origin and midpoint of the membrane was set to a water molecule in the center of the channel, which corresponds to W522 in the X-ray structure [88]. The membrane, which was described using a dielectric model, was set to have a thickness of 35 Å; the dielectric coefficients of membrane and bulk solvent were chosen to be 2.0 and 80.0, respectively, and the salt concentration

was chosen as 150 mM. Similar to the model channel simulations discussed above (sec. 4.2), Legendre polynomials up to order 10 were used as basis functions for the GSBP inner region electrostatics. The protein-solvent boundary was set up using the atomic Born radii of Nina et al. [298]. The atoms outside the inner cavity were constrained to their initial position. Protein atoms at the boundary of inner and outer cavity were constrained according to the previously described protocol[135]. For the Coulombic interactions, the extended electrostatics model [223] was used where interactions beyond 12 Å were treated using a multipolar expansions including dipole and quadrupole terms.

The model contains a total of 154 water molecules, 27 of which were treated using SCC-DFTB. Treating these water molecules with a QM model provides a consistent and natural description for the PT process; the cooperativity among the water molecules, which is missing in previous classical treatments [88, 89, 105, 106, 110], is also taken into account. The validity of the QM/MM model for water has been shown using model channels in section 4.2. The comparison of the PT barriers for the QM/MM scheme with a full QM approach indicated that solvation is adequately described using MM water molecules in a QM/MM scheme.

All of these QM waters were in the interior of aquaporin in the starting configuration. To prevent lateral diffusion of QM and MM waters, planar restraints were introduced at the interface between QM and MM water molecules. The QM/MM frontier and respective restraints were chosen such that additional QM water molecules can diffuse into and out of the pore. The reservoirs of QM water (red) can be seen on each side of the monofilar water chain in Fig. 5.7. This goes beyond the setup in a previous study [106], which reproduced the location of the water molecules in the channel but was limited in its predictivity by restricting water from entering or exiting the pore.

The QM/MM model including parts of the protein backbone in the QM segment additionally comprises the backbone atoms of the residues 63 to 69 and 198 to 204. The boundary between the QM and MM region was chosen at the C-C_α bond such that the carbonyl groups of residue 63 and 198 are included and those of residue 69 and 204 are excluded from the QM segment and at the C_α-C_β bond of the residues 65 to 67 and 200 to 202. Link atoms were introduced to saturate the valence of the QM boundary atoms. The interactions at the boundary are treated using the EXGR scheme implemented in CHARMM. The validity of the partitioning across the C-C_α bond in the backbone was verified by comparison of the Mulliken charges of the QM segment from a QM/MM calculation of the functional loops with results from calculations treating the complete functional loops with SCC-DFTB.

The GSBP-QM/MM dynamics simulations employed constant-temperature MD using the weak coupling method for temperature control described by Berendsen et al. [299] with a time step of 1.0 fs and a temperature of 300 K.

For the simulation of the water structure, eight trajectories were calculated, each with a time period of 20 ps for equilibration and 100 ps of data collection. For the determination of relative water densities, the pore radius of the lumen was determined at each point along the channel axis using the Hole2 program [300].

In addition, a set of calculations without the dielectric model of membrane and bulk solvent, i.e., the environment of the protein was replaced by vacuum, was also car-

ried out for comparison with the GSBP results. In the following these these calculations are referred to as vacuum simulations. The length of the MD simulations was the same as in the GSBP based simulations.

PMF calculations were carried out using the standard umbrella sampling technique [273] with the modified center of excess charge coordinate ξ_z [130] (see section 4.1.2) as the reaction coordinate since the water wire in GlpF is linear. The umbrella sampling involved the simulation of 57 independent windows with force constants of 10 kcal/mol Å^2. The simulations each consisted of 20 ps of heating and equilibration, and a further 80 ps for production and data accumulation. To obtain the PMF from the accumulated data, the WHAM equations were employed.

Despite the obvious advantages of PMF computations, the methodologies of analysis lag behind the ones routinely used for minimum energy pathways (MEP) or transition state (TS) computations. In TS or MEP computations the perturbational analysis of the energetics by recomputing single point energies at the critical points of the potential energy surface (PES) with an alternative functional belongs to the standard protocol [301]. This technique allows for an estimate of the influence of different groups or residues using *in-silico* mutations and allows the explorations of factors influencing the energetics of the reaction *without* recomputing the MEP.

Similar techniques can be used to investigate free energy integration computations using force decomposition or perturbational techniques [302]. For umbrella sampling, changes in the forces in direction of the reaction coordinate can be used to perturbationally explore the change in the potential of mean force upon alteration of the potential energy functional [303]: Using structures recorded during the umbrella sampling simulations, the change in the forces in direction of the reaction coordinate can be computed. Integration yields the perturbational change in the PMF. Employing this technique, the energetical influence of specific elements was investigated by using charge-modified variants of the channel (Table 5.1).

5.3.2 Water Structure in GlpF

The water structure has been proposed to contribute to the high barriers for PT in GlpF and other aquaporins proteins [88–90, 105]. Although recent simulations of aquaporins have argued against an explanation of the proton blockage based on the structure of the water chain, aquaporin water channels have been shown both experimentally [94] and through molecular dynamics simulations [88, 89, 105, 106, 110] to have a pronounced water structure. Similar observations were also made in gramicidin in simulation and experiment as reviewed in ref [304].

The embedded water structure has been confirmed by multiple simulations using different force fields (GROMACS, CHARMM), water models (TIP3P, SPC, PM6) as well as different simulation protocols [88, 89, 105, 106, 110]. For example, Tajkhorshid et al. [88] carried out simulations for GlpF: a tetramer was embedded in an explicit membrane-water environment and the Ewald summation for long-range electrostatics was applied with periodic boundary conditions. The simulations showed distinct water structure around hydrogen-bond-forming moieties in the lumen region (see Fig. 5.5), and the locations of high water densities match well with those identified from high-resolution X-ray structures presented in the same paper.

Comparing results from GSBP-QM/MM simulation to those previous calculations using explicit membrane and bulk solvent molecules is expected to be a stringent test of the used simulation protocol.

Molecular Dynamics Simulations of the Water Structure

As shown in Fig. 5.8 A, the distribution of water molecules has distinct features inside the aquaporin. The peaks in the relative water distribution coincide well with the position of specific protein polar groups. These include the well-conserved NPA motif in the center of the channel and two half membrane spanning, short peptide sequences (65-68, 199-203) whose carbonyl groups line the channel; Arg206 in the selectivity filter region also helps to position water molecules. The role of Trp48 and Phe200 is mainly attributed to narrowing the pore size and providing the hydrophobic corner discussed above. This might improve the interaction of the water molecules with Arg206 [89]. These observations are in good agreement with previously published data [88, 89] using models that are substantially larger in size. Moreover, the bifurcating orientation of the water wire centered around Asn203 as observed in previous studies is also reproduced by the current simulation (Fig. 5.8 B). The order parameter $P_1(z) = \langle \cos\theta \rangle$ with θ as the angle between the membrane normal and the water dipole vector shows a sigmoidal shape with an inflection point at z≅-2 Å, i.e., at the position of Asn203, reflecting the known [88, 105] bipolar water orientation in the cytoplasmic and periplasmic half of the channel with the inversion of the water dipole orientation at the NPA motifs.

In the additionally conducted vacuum simulations, the distribution of water agreed well with GSBP only in the central region of the channel. The agreement is visibly worse in regions beyond 5 Å from the center. In particular, the peak positions of the water molecules are shifted in the region of negative z value; these shifts can not be attributed to the shifts in the position of polar groups in the lumen, which are displaced by less than 0.1 Å in the vacuum simulations in comparison to the GSBP simulation. Moreover, disruptions of the water chain were observed between z= 5-10 Å, which is clearly visible in the average density profile. Previous periodic boundary based simulations [105] did notice disruption of the water wire inside aquaporin, although these events were very rare. Therefore, the observed breakdown of the well-ordered water structure in the vacuum simulation is unexpected for a channel with many polar interactions despite its narrow radius. In previous work, such pulsatory behavior was observed only in narrow and *purely hydrophobic* channels, such as a hydrophobic model of the aquaporin channel [105], carbon nanotube [305], and smooth hydrophobic pore [306] and cylinders [307]. In fact, since the pulsatory transport of water in narrow pores depends on collective oscillation of the bulk water [306], the lack of large body of bulk solvent in the current set-up is not expected to produce channel depletion.

The reproduction of the water density from experiment [88] and simulations with an explicit description of the larger environment [88, 89] confirms the simulation protocol for the following free energy calculations of the PT reaction through GlpF.

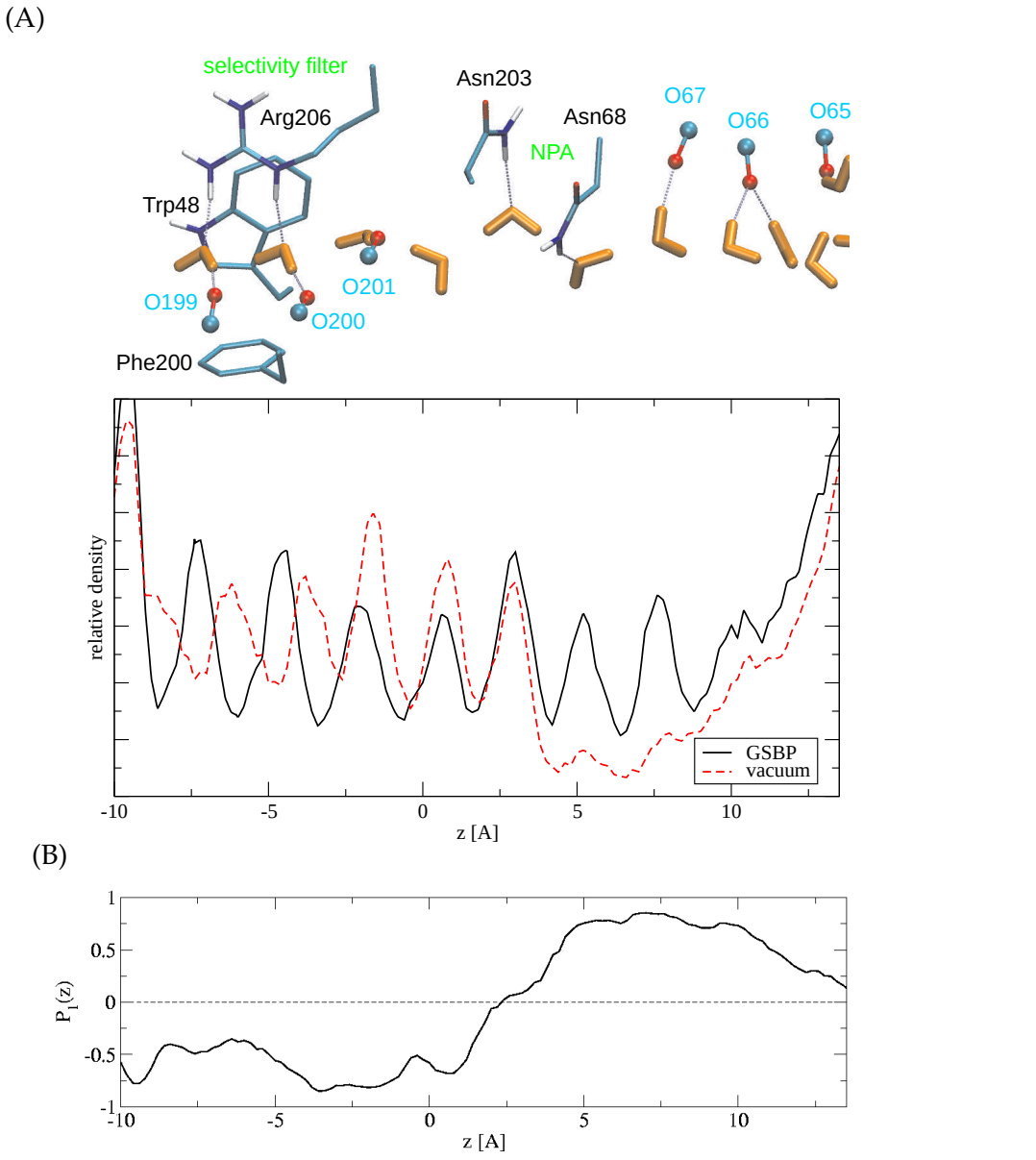


Figure 5.8: (A) Computed relative water distribution in the GlpF channel along the membrane norm (z axis) for the vacuum- and the GSBP-based simulations, which were obtained through the frequency of observing the water oxygen atoms at a the respective position. The average positions for selected atoms from the GSBP simulations are also shown to illustrate the correlation between water density and position of polar groups along the channel, which include both protein side chains (e.g., Arg206, Asn68, Asn203) and main chains (e.g., a series of carbonyl groups). (B) Water orientation in the GlpF channel illustrated by the order parameter $P_1(z) = \langle \cos\theta \rangle$; θ is the angle between the membrane normal and the normalized water dipole vector.

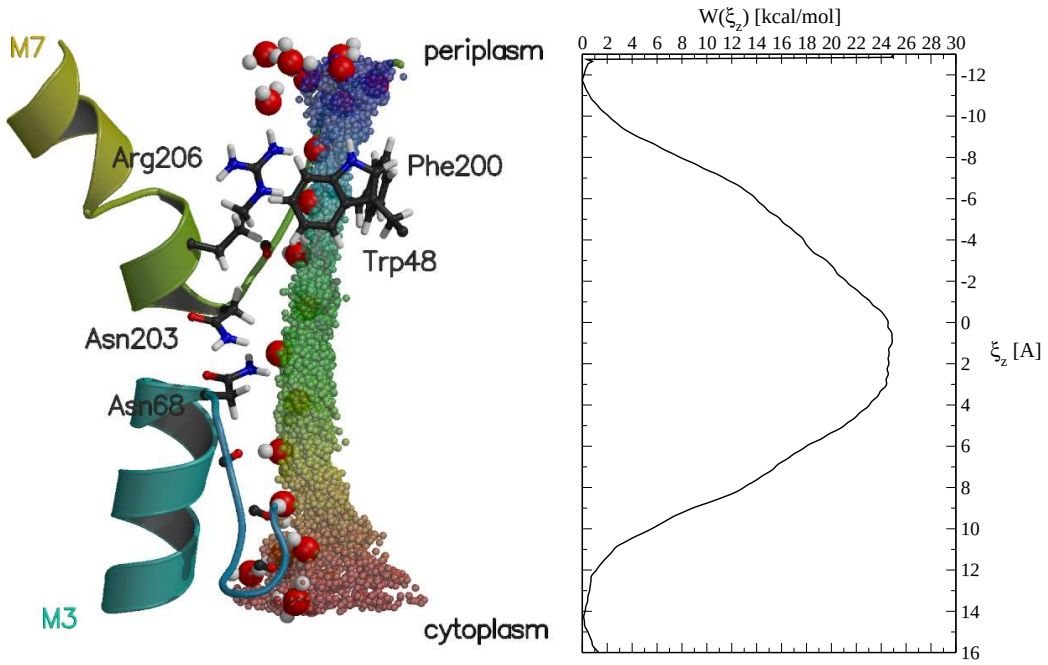


Figure 5.9: (A) Trace of the location of the excess proton calculated using the mCEC. The color displays the value of the ξ_z coordinate from -12.0 Å (blue) to 16.0 Å (red). (B) PMF for the proton transfer through GlpF.

5.3.3 Potential of Mean Force Calculations

The PMF profile for the PT through wild-type GlpF shows a barrier of 24.9 kcal/mol peaking in the middle of the channel/membrane (ξ_z=0.0 - 3.0 Å) close to the NPA motifs. The profile further features a shoulder at the position of the selectivity filter at $\xi_z =\sim$ -6.0 Å (see Fig. 5.9). During the PT reaction, also the water orientation in the pore was calculated using the order parameter P_1 (Fig. 5.10). The sigmoidal shape of the profiles indicate bipolar water orientations whose center gradually change during the PT reaction depending on the reference value of the umbrella sampling biasing potential. The center of the bipolar water orientation is therefore determined as expected by the position of the excess proton (see Fig. 5.10).

To gauge the influence of specific elements on the PT energetics, nine systems denoted A to I (see Table 5.1) with modified charge distribution of the channel and dielectric properties of its environment were analyzed. For each modified system, the respective energy terms of a single or more elements were set to zero and the perturbational correction to the PMF calculated. The perturbations A to E thereby concern localized *structural* elements in the pore, whereas the perturbations F to I concern *global* aspects. The perturbations only enter energetically without altering phase space, in contrast to approaches which already turn off elements during the simulation [107, 112]. Hence, the importance of each element on the actual PT can be separated from structural and dynamical effects on the protein and the water file through it.

Turning off the partial charges on the side chain of Arg206 (model A) causes a sig-

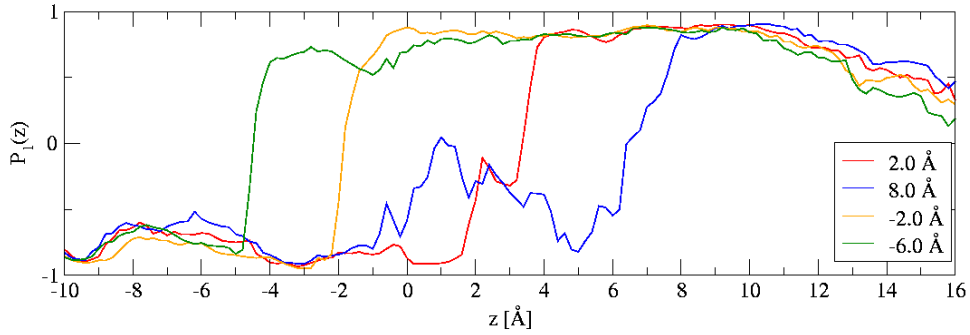


Figure 5.10: Water orientation in the GlpF channel for different reference values of the umbrella sampling biasing potential illustrated by the order parameter $P_1(z) = \langle \cos\theta \rangle$; θ is the angle between the membrane normal and the normalized water dipole vector.

nificant negative perturbational correction to the PMF at the selectivity filter region. Therefore, Arg206 significantly increases the free energy barrier in this region. The maximum correction of -6.4 kcal/mol is reached at ξ_z=-5.6 Å which is close the position of the C_ζ of Arg206 along the channel axis (Fig. 5.11 A). To larger values of ξ_z, the negative correction is continuously reduced, vanishes at the NPA motifs, and increases to positive values (up to 2 kcal/mol) in the cytoplasmic half of the pore.

The perturbations of the models B, D, and E have only a minor influence on the PMF and the barrier height. Turning off the partial charges of NPA asparagine side chain atoms (model B) results in a localized negative correction of 1.6 kcal/mol at the NPA region (ξ_z= 3.0 Å). To larger and smaller values of ξ_z, the negative correction declines quickly and reaches small positive values ($\sim$ 0.7 kcal/mol) at the selectivity filter region and in the second half of the cytoplasmic half of the pore.

The perturbation in model D (backbone atoms of the M3/M7 helices) causes a maximal correction of -3.0 kcal/mol at ξ_z=3.6 Å. However in contrast to model B, a broad minimum covering the complete channel is obtained. A similar result is found for the QM treatment of the functional loops between the helices M2/M3 and M6/M7 (model E). The perturbation leads to a negative correction for almost the complete pore with two local minima of -2.1 kcal/mol and -2.0 kcal/mol at ξ_z=-2.3 Å and 7.6 Å.

Inversely to model D, the C=O groups lining the channel (model C) cause a significant stabilization of the excess proton in the channel. $\Delta W(\xi_z)$ exhibits an inverted double-well profile with maxima of 6.3 and 5.2 kcal/mol located in the middle of cytoplasmic and periplasmic half of the channel at ξ_z=-2.6 Å and 8.1 Å. A local minimum with a value of 3.8 kcal/mol is found at the NPA motifs region.

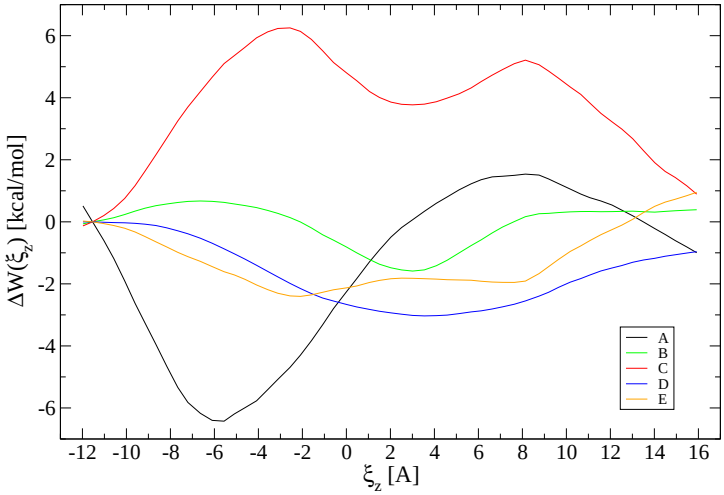
For model F ("hydrophobic" channel), the perturbational correction to the PMF shows a minimum of -2.5 kcal/mol at the NPA region (ξ_z=2.5 Å) (Fig. 5.11 B). To larger and smaller values of ξ_z, the negative correction declines quickly and becomes positive with two local maxima of 3.0 and 2.5 kcal/mol in the cytoplasmic and periplasmic half of the channel at ξ_z=-4.0 and 8.8 Å. A very similar results is obtained if additionally to the point charges of the protein the membrane and the

Table 5.1: Perturbations of the GlpF channel studied. [a]

	model								
perturbation	A	B	C	D	E	F	G	H	I
Arg206[b]	-								
NPA(Asn68/203)[b]		-							
C=O[c,d]			-						
M3/M7 helix backbone[c,e]				-					
M2/M3 & M6/M7 loops backbone QM[f]					-				
protein[c]						-	-	-	
membrane/bulk solvent[g]							-	-	
MM water[c]								-	-

[a] The minus sign indicates that the respective energy contribution is turned off. [b] Partial charges of the side chain atoms turned off. [c] Partial charges turned off. [d] Residues 64-66 and 199-201. [e] Residues 70-79 and 205-217. [f] Backbone of residues 63-69 and 198-204 treated quantummechanically. [g] GSBP turned off.

(A)



(B)

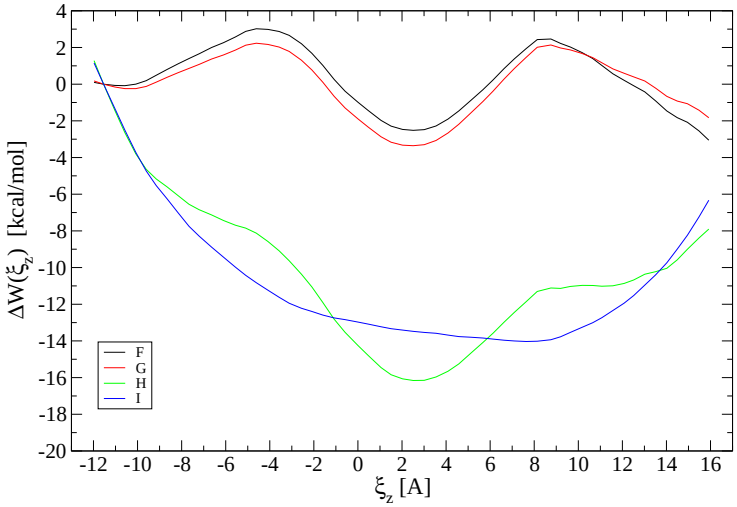


Figure 5.11: Perturbational correction to the PMF for localized structural elements(A) and global aspects (B).

bulk solvent (model G) are turned off. The minimum at the NPA region is slightly deeper (-3.4 kcal/mol) and the maxima are slightly lower (2.2 and 2.1 kcal/mol).

In contrast, turning off the partial charges on the MM water molecules, either for the "hydrophobic" channel with neglected membrane/bulk solvent (model H) or solely (model I) causes a huge negative correction to the PMF around the NPA region (Fig. 5.11 B). For model I, a broad minimum spanning the whole channel is found with a maximal correction of -14 kcal/mol. The perturbational correction for model H shows a more structured profile with a distinct minima of -16.2 kcal/mol at the NPA motifs (ξ_z=2.5 Å).

5.4 Discussion

Several theoretical studies [88–90, 104–112] have addressed the mechanisms of the water/proton selectivity in aquaporins. Despite this effort, the exact nature of the mechanism is still disputed. In part, the discrepancies also with regard to quantitative differences in the barrier heights originate in the wide variety of used computational methods and simulation schemes ranging from classical electrostatic calculations [105, 106, 108] to an ab initio description of the water wire [297]. Thus it remains meaningful to carry out reliable QM/MM simulations to better understand the importance of various factors that have been proposed.

The presented SCC-DFTB/CHARMM/GSBP set-up uses a fairly small model for aquaporin ($22\times25\times54$ Å^3, which contains ∼2500 atoms) and treated both bulk solvent and the lipid membrane using continuum models. The simulation reproduced the water structure and bipolar water orientation in the GlpF channel compared to previous periodic boundary simulations with explicit lipid membrane and bulk solvent [88, 89] and X-ray structures [88]. This is quite remarkable considering the much smaller number of atoms explicitly included in the present protocol and the fact that the membrane was replaced by a simple dielectric model. Simulations of the same reduced system with the protein surrounded by vacuum, by contrast, produced rather different water distributions, which further illustrates the importance of proper long-range electrostatic treatment for intrinsically heterogeneous systems such as membrane-bound proteins and the effectiveness of the GSBP approach. Furthermore, the results underline the validity of the model to simulate PT reactions.

The computed value for the barrier height of ∼25 kcal/mol for the PT through GlpF is large enough to prevent the leakage of protons into the cell. Similar results have been obtained in previous studies – 28 and 26 kcal/mol for AQP1 [111, 112] and 18 kcal/mol for GlpF [110]. Furthermore, the free energy values for the barrier height are close to values found for the transport of a proton including the formation of membrane spanning water networks around the excess proton (18 kcal/mol) [308] and the formation of a water file through a complete membrane (26 kcal/mol) [309]. This supports the previously brought up argument that there is no need for aquaporins to oppose PT stronger than the surrounding membrane itself [108]. The clear maximum of the PMF at the NPA motifs region, however, does not allow a conclusion regarding the nature of the barrier since the NPA motifs region is also simply the center of the low dielectric region of channel and membrane.

The QM treatment of the backbone of the functional loops (model E in Table 5.1) results in the expected stabilization of the excess proton in the channel as qualitatively also shown by Jensen et al. [105] in their analysis of the electrostatic interaction energies of a positive point charge with the protein. It allows to estimate the error of the barrier height due to the neglect of protein polarization and thus to asses the validity of the employed QM/MM approach using a nonpolarizable force field. The amplitude of the found correction to the PMF is small (∼2 kcal/mol) in contrast to the results found for carbon nanotube channel in which wall polarization effects were found to be significant [310]. Therefore, the neglect of protein polarization in the computational model is not expected to alter the results of the investigation critically and the application of a QM/MM scheme using a nonpolarizable force field is justified.

In addition to the PMF for the PT that provides only limited information about the mechanism of the water/proton selectivity, the results of a perturbation analysis provide a detailed picture of the contributions to the barrier. The investigated perturbations (Table 5.1) can be separated into perturbations concerning localized structural or global elements as, for example, Arg206 (model A), the NPA motifs (model B), the helices M3/M7 (model D) or the protein as a whole (model F) and and the remaining perturbations which describe a change of the reference point of the PT reaction through the channel (model H and I). In contrast to previous approaches [107, 112], the employed perturbation analysis thereby considers only the energetics of specific structural elements to the PMF, but preserves the structure of the water file through the pore and the protein structure. Therefore, it allows to separate the effects due to conformational changes such as an altered water structure in the pore from those due to the electrostatic influence of structural elements.

Among the structural elements, the NPA motifs (model B) and the helix macrodipoles (model D) do not contribute much to the free energy barrier for the PT through GlpF. Turning off the partial charges of Asn68 and Asn203 (model B) results only in a correction to the PMF of -1.5 kcal/mol. This agrees with the small contribution of the NPA motifs to the electric field in direction of the channel axis at the position of the water file found in a previous analysis [105]. Similar contributions to the barrier were found in previous free energy simulations (∼3 kcal/mol [107] and 3.6 kcal/mol [111]) and electrostatic energy simulations (∼2 kcal/mol [107]). Nevertheless, the NPA motifs play an important role for the stabilization of the bipolar water orientation of the nonprotonated water chain through GlpF [88, 105, 107]. Against a possible secondary influence on the PT due to the stabilization of the bipolar water orientation, however, argues that the water orientation is a dynamical phenomenon. It can readily disrupted by thermal fluctuation as Miloshevsky and co-workers [104] in their MC study showed. In line with this are the results for the water orientation in the pore during the PT (Fig. 5.10). The results show that the original bipolar water orientation is readily adapted to the location of the excess proton.

Similar to NPA motifs, the small correction of -3 kcal/mol for the helices M3 and M7 (model D) indicates only a minor contribution to the free energy barrier. In general, the influence of these helix macrodipoles is a disputed topic. Sengupta et al. [311] pointed out that the net effective dipole moment of a helix in proteins can in fact be of considerable amount. However, instead of examining the net dipole moment of a helix its contribution to the electric field at the points of interest is more meaningful.

So, Åqvist et al. [312] showed that the stabilizing/destabilizing effect of a helix results rather from some localized dipoles located in the first turn of the helix than from the macrodipole of the helix and is therefore smaller than usually assumed.

As for the NPA motifs, it was shown that the contribution of the M3/M7 helix to the electric field is vanishingly small [105] suggesting only a small contribution to the energetics of PT as found here. Similar contributions to the barrier (∼2.5 to 4 kcal/mol) were found in previous free energy and electrostatic energy simulations [107]. A most recent study [112], however, found a significant larger influence (16 kcal/mol). Though, a comparison of the result is hampered since the perturbations in this work only enter energetically and do not influence the structure of the water file through the pore. In ref [112], in contrast, the electrostatic effect of the helices and structural changes of the water file, as observed, are mixed. To what extend the structural changes influence the PT remained unclear.

In contrast to the NPA motifs and the helix macrodipoles, the positively charged Arg206 of the SF (model A) contributes significantly to the the PMF. However, a large correction (-6 kcal/mol) to the PMF is found only at the SF whereas the correction at the barrier maximum is about zero. Therefore, the shoulder in the PMF is affected while the height of the free energy barrier itself and the proton blockage remains unaffected, even if the shape of the PMF becomes asymmetrical as also observed in previous studies [107, 112]. Unexpectedly, a recent experimental study indicated that after the mutation of the Arg (Arg195) in the SF of AQP1 to Val, AQP1 allows the passage of protons [102]. Inspired by this, the influence of the SF on the proton blockage in AQP1 was investigated in a theoretical study, and a decrease of the free energy barrier from 28 kcal/mol to ∼16 kcal/mol if Arg195 is mutated to Val [112] was found. The obtained barrier height for the mutant, however, is still too high to allow PT.

In both studies [102, 112], the found effects reflect two interfering sources - on the one hand, the changed electrostatic interaction between the positively charged arginine and neutral valine with the excess proton and, on the other hand, the change of the structure at the SF region. Mutating Arg to Val widens the narrow SF [102] and may cause additional water molecules to fill the additional space and thus provide a better solvation for the excess proton. A X-ray structure of the mutants, however, is not available to verify this aspect. Also the second mutation in ref [112] supports this view. The mutation of the neutral His180 to Ala further widens the SF, accordingly, an additional drop of the barrier is found [112]. However, the electrostatic interaction of the neutral His with the excess proton can be assumed to be of minor importance according to ,e.g., the results for the Asn side chains of the NPA motifs (model B). Hence, further studies are needed to quantify the contribution of the SF to the blockage of protons by establishing a narrow constriction region with the therewith connected desolvation effects of the excess proton. The energetical influence of the Arg, by contrast, is small as the results for model A indicate.

In addition to these localized perturbations, the influence of the protein as a whole is investigated in model F. Turning off all partial charges of the protein leaves a "van-der-Waals channel" that shows almost the same free energy barrier like the wild-type channel as the small correction to the PMF indicate. This results indicates that the electrostatic interaction of the protein matrix with the excess proton does **not**

dominate the free energy barrier. A similar results is obtained if additionally the continuum electrostatic description of the membrane and bulk solvent are turned off. These results agree with results by Burykin et al. [109] who obtained similar free energy profiles for the PT through AQP1 and a hydrophobic model channel.

In contrast to the perturbations above, the MM water molecules (model H and I) significantly contribute to the free energy barrier. Turning off the partial charges of the MM water molecules reduces the solvation of the excess proton at the end of the single water file and thus destabilizes the excess proton. As result an enormous negative contribution (-16.0 kcal/mol for model H and -14.0 kcal/mol for model I) to the free energy barrier is found. This result indicates that the major contribution to the free energy barrier results from the desolvation penalty for transferring a proton from the bulk to the single water file. In other words, turning off the partial charges of the MM water molecules changes the reference point of the PT through the pore. The remaining low barrier, therefore, represents the contribution of the protein alone with solvation effects due to bulk solvent eliminated. This result agrees very well with the results obtained for the model channel embedded in various environments in section 4.2. The different environments cause a different preferential solvation at the ends of the water wire and thus modulated the barrier height for the PT.

Summarizing, the results indicate that the major contribution to the free energy barrier results from desolvation effects of the excess proton. Structural motifs such as the NPA motifs are of minor importance but of course influence the barrier. Polar groups in the pore, on the other hand, stabilize the excess proton in the pore as the results for the carbonyl groups of the two functional loops (model C) indicate. However, in aquaporins their effect is not large enough to counterbalance the penalty due to the desolvation as for instance in gramicidin. Its pore exhibits a stack of sequential rings of carbonyl oxygen atoms [91, 92] that compensate the energetical cost of dehydration. A similar situation was investigated for the model channel in section 4.2. Circular arranged dipole moments around the model channel were shown to modulate the PT barrier significantly resulting in a nearly vanishing barrier.

An interesting question is how these results behave to the recent experimental study that found proton conduction in AQP1 for the mutants of the Arg of the SF [102]. Note that for such mutations two effects have to be separated. The analysis in this work considers the electrostatic effect due to a mutation – an experimental mutation on the other hand involves an electrostatic effect and an influences on the structure. Mutation of the residues constituting the selectivity filter most likely increase the cross section of this constriction region and additional water molecules increase the solvation of an excess proton decreasing the barrier height.

To investigate this issue, further experimental and theoretical studies are necessary. On the one hand, a X-ray structure would provide the required information about the structural alterations of the protein matrix and the water structure at the SF region. Theoretical methods, on the other hand, can, as complementary method, contribute by simulating the water distribution in the pore region, which is often difficult to access by X-ray diffraction.

Chapter 6

Color Tuning in Archaeal Rhodopsins: Bacteriorhodopsin vs Phoborhodopsin

The discovery of bacteriorhodopsin (bR) [7, 10, 313–315] in the seventies [316] as a proton pump in Halobacteria which is not in direct contact with energy-transducing proteins provided a possibility to explore the chemiosmotic hypothesis proposed at that time. Furthermore, together with other halobacterial rhodopsins it set the stage for the investigation of the molecular mechanism of the color tuning in rhodopsins. Apart from bR, three more rhodopsins are found in the cellular membrane of Halobacteria: phoborhodopsin (pR; also called sensory rhodopsin II, sRII) [10, 317, 318], halorhodopsin (hR) [319], and sensory rhodopsin I (sRI) [10, 317, 318].

The blue shifted absorption maximum (λ_{max}) of pR ($\sim$500 nm) relative to bR, hR, and sRI, which absorb in the range of 560-590 nm, is very remarkable since halobacterial rhodopsins share the same chromophore and are highly homologous in their structure. The amino acid sequence of the phoborhodopsin is $\sim$25 % identical to that of bR [320]. If only the residues that constitute the binding pocket are considered, the identity increases to $\sim$80 %; only 10 amino acids within 5 Å of the chromophore [320, 321] differ. Similar degrees of conservation are found for the binding pockets of sRI and hR [10]. Because of this ostensible conflict, the spectral shift between bR and ppR has been studied extensively with experimental [27, 322–325] as well as theoretical methods [50, 51]. The clarification of the molecular mechanism of the shift also sheds light onto the general mechanisms of color tuning in retinal proteins and is therefore of fundamental importance for understanding the process of color vision in vertebrates where homologous rhodopsins are found.

In the next two sections 6.1 and 6.2, the structure and photocycle of bacteriorhodopsin and phoborhodopsin are shortly reviewed, followed by an outline of the current state of knowledge on color tuning focused on the investigated systems (sec. 6.3). In section 6.4 the computational strategy is introduced and the applicability of various computational methods is discussed. Subsequently, the mechanisms

of the color regulation between bR and ppR are investigated in section 6.5. A summary of the results finally concludes the chapter (sec. 6.6).

6.1 Structure and Photocycle of Bacteriorhodopsin

Structure

The structure of bR is one of the most highly resolved amongst membrane proteins. First structural models had been determined by electron diffraction methods at increasing resolutions to 3 Å [326, 327], but first the development of the lipidic cubic phase crystallization [328] allowed for preparation of high-quality crystals for atomic-resolution X-ray diffraction. Since then the resolution has improved significantly to recently 1.43 Å [329–333] providing information about the precise location of amino acid side chains and internal water molecules.

In the membrane bR clusters into trimers [5] (Fig. 6.1) that form a two-dimensional hexagonal lattice, the purple membrane. Under low light conditions, the purple membrane contains an equilibrated mixture of bR containing all-*trans* and 13-*cis* conformers [334] of retinal *. However, only the protein with the all-*trans* chromophore is biologically active (proton pumping); the 13-*cis* protein also absorbs light but transforms into the all-*trans* protein, a process referred to as light adaption [233, 335, 336].

Bacteriorhodopsin consists of 248 amino acids folding into a seven-transmembrane helical topology (Fig. 6.2) [326]. The helices (A to G) are connected by short loops and grouped into two layers of three and four helices. The latter forming an arc-like structure (Fig. 6.1). All helices are oriented perpendicular or nearly perpendicular to the membrane. Retinal is connected via a Schiff base linkage to Lys216 on helix G and separates the protein into a cytoplasmic and an extracellular half. The plane of the chromophore is oriented nearly perpendicular to the membrane plane [337] and its β-ionone ring end is tilted toward the extracellular side [338]. In the light adapted ground state, the NH group of the protonated Schiff base (PSB) is oriented toward the extracellular side [339]. The β-ionone ring/chain conformation is 6-s-*trans*.

Surrounding retinal, three tryptophanes (Trp86, 182, 189) and a tyrosine residue (Tyr185) tightly pack the chromophore and fix the polyene chain in the proper orientation and position [340]. The environment around the PSB is formed by Thr89 and the counterion residues Asp85, Asp212, and Arg82 (Fig. 6.4) which stabilize the positive charge of the PSB. Moreover, the Schiff base region contains several water molecules [330, 332]: a highly polarized water molecule coordinated to the PSB, Asp85, and Asp212 as well as two water molecules coordinated to Asp85 and Asp212, respectively. Together with the counterion residues, they constitute a hydrogen bonded network (HBN) (Fig. 6.4).

A second HBN of water molecules in the extracellular half is separated by Arg82 from the Schiff base region (Fig. 6.2). Together with the counterion residues and several polar side chains †, the water molecules in the extracellular half constitute

*See Fig. 6.3 for numbering of retinal atoms.

†Thr89, Tyr83, Tyr57, Tyr185, Glu194, Glu204, Ser193

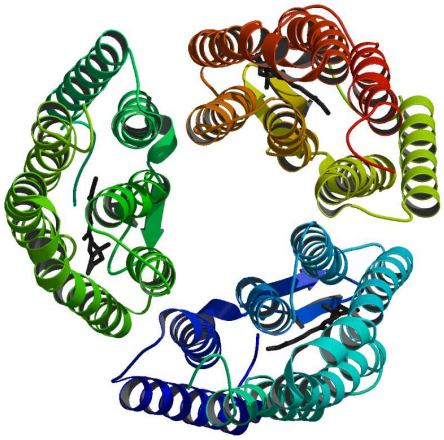


Figure 6.1: Structure of the bacteriorhodopsin trimer in the purple membrane.

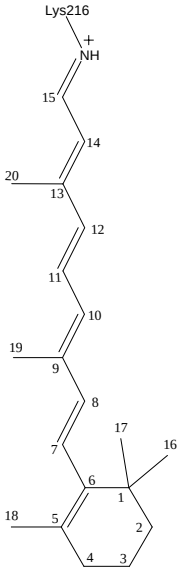


Figure 6.3: Chemical structure of the retinal chromophore in the all-*trans* configuration.

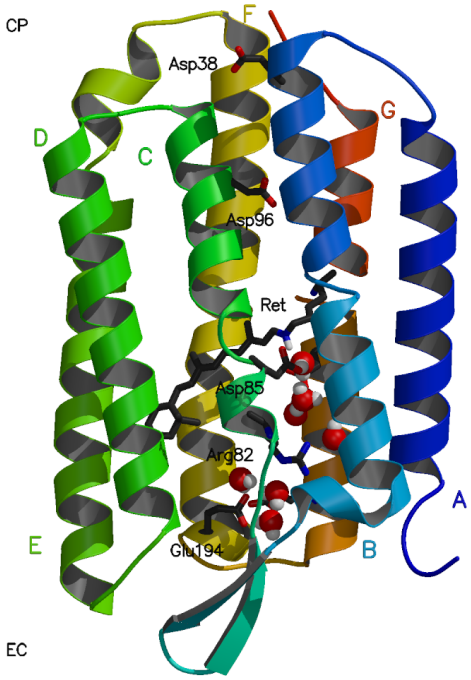


Figure 6.2: Structure of bacteriorhodopsin (PDB code 1c3w) with the chromophore and selected functional important residues highlighted in ball-and-stick representation. Selected water molecules are shown in CPK representation.

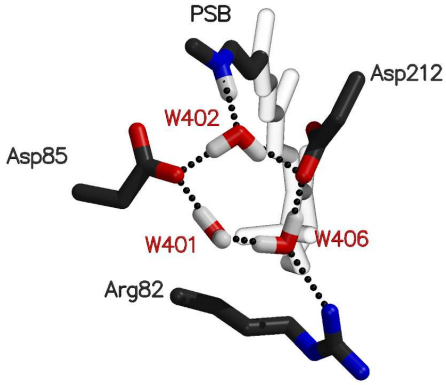


Figure 6.4: Detailed view of the retinal Schiff base region of bR (PDB code: 1c3w) showing the retinal and its counterion residues Asp85, Asp212 and Arg82 as well as the three water molecules of the hydrogen-bonded network.

a three dimensional network reaching from the PSB to the surface [315] that way constituting a proton transfer pathway.

Interestingly, there is no obvious proton conduction pathway in the cytoplasmic half. The amino acids Asp96 and Asp38 were found to be involved in the transport [341–343], but there are hydrophobic gaps, for instance between Asp96 and the Schiff base (12 Å), which have to be bridged. While X-ray structures could not verify a water chain directly despite several cavities [326], FTIR spectroscopy [344] and molecular dynamic simulations [345, 346] suggested hydrogen bonded networks in the cytoplasmic half. Therefore, transiently formed aqueous chains were postulated.

Photocycle

The absorption of light leads to photoisomerization of the chromophore from the all-*trans* to a 13-*cis* conformation and triggers the following commonly accepted photocycle. The subscripts denote the wavelength of the visible absorption λ_{max}; the approximate transition lifetimes are shown above the arrows [7, 347–349]:

$$bR_{568} \overset{h\nu}{\rightsquigarrow} I_{480} \overset{\sim 0.5\ ps}{\longrightarrow} J_{625} \overset{\sim 2\ ps}{\longrightarrow} K_{610} \overset{\sim 2\ \mu s}{\longrightarrow} L_{550} \overset{\sim 80\ \mu s}{\longrightarrow} M_{412} \overset{\sim 10\ ms}{\longrightarrow} N_{520} \overset{ms}{\longrightarrow} O_{640} \rightarrow bR_{568}$$

Formally, six steps consisting of isomerization, proton transport, and a switch of the accessibility of the Schiff base characterize the photocycle. After the isomerization of retinal during the bR→J transition, following structural relaxations [329, 350–355] set the stage for the first PT step from the Schiff base to Asp85 during the L→M transition [356]. During the lifetime of M, a proton (second PT step) is released to the extracellular aqueous phase from a group close to the extracellular surface (proton release group PRG) [357–359], whose nature is still disputed [333, 360–363]. Further, the accessibility of the Schiff base switches from extracellular to intracellular to allow vectorial proton transport [364–366]. The Schiff base is then reprotonated from Asp96 (third PT step) in the fourth step during the rise of N [342, 343, 356]. During the following decay of N [342, 343, 367], Asp96 is reprotonated from the cytoplasmic side and retinal reisomerizes thermally [368]. In the last step, a PT from Asp85 to the PRG [342, 367, 369] and structural relaxation to the initial state [370] during the decay of O complete the photocycle.

6.2 Structure and Photocycle of *pharaonis* Phoborhodopsin

Besides in *Halobacterium salinarum*, phoborhodopsin is also found in *Natronobacterium pharaonis* [371]. But X-ray structures are only available for the *pharaonis* phoborhodopsin (ppR, also psRII), wherefore in this work ppR is investigated. The absorption spectra of ppR and *salinarum*pR (spR or ssRII) are very similar showing only a shift of 10 nm of the absorption maximum λ_{max} (497 nm for ppR [20, 372] and 487 nm for spR [373]).

Structure

In the membrane, ppR is complexed with a transducer protein (HtrII) which binds specifically [374] by means of two transmembrane helices (TM1 and TM2) and links the photoreceptor ppR with a cytoplasmic signal cascade. Two of those receptor/transducer complexes in turn cluster to a dimer with a two-fold rotation axis [375, 376] (Fig. 6.5). High resolution crystal structures of the receptor/transducer complex [376, 377] and transducer-free ppR [320, 378] show that the structure of ppR is almost identical in both cases. However, functionally transducer-free ppR exhibits light-induced proton pump activity [379–381] which is first inhibited with the complexation of the transducer [381–383].

The protein consists of 239 amino acids and is 27 % identical to bR in amino acid sequence [320, 384]. The general helical structure is similar to that of bR (Fig. 6.6). The helices C to G show the highest similarity to bR with a rms deviation of the backbone atoms of only 0.77 Å [378]. Regarding the interconnecting loops, the antiparallel β-sheet at the BC loop is shorter in ppR than in bR [320] (Fig. 6.6).

Retinal is bound to Lys205 on helix G and its conformation is exclusively all-*trans* [385] in the ground state. In contrast to bR, there is no light-dark adaption [386]. The chromophore is mainly unbent (Fig. 6.6); the β-ionone ring is in a 6-s-*trans* conformation as in bR, and the NH group of the PSB points to the extracellular side.

A comparison of the binding pockets of ppR and bR displays that most residues are conserved. Only ten residues in 5 Å range around the chromophore are different [320, 321]. Most interestingly are the replacements of Gly130Ser, Ala131Thr, and Thr204Ala [‡] which alter the polarity of the binding pocket and the replacement of Val108Met which together with Gly130Ser modifies the steric environment around the β-ionone ring.

In the Schiff base region, Asp75, Asp201, and Arg72 like their counterparts in bR form the complex counterion of the PSB, but unlike to bR, Arg72 points toward the extracellular side (Fig. 6.7). Just as in bR, the Schiff base region contains three water molecules (W402, W400, W401) [320, 378] forming a HBN. Additional water molecules are located in a second cluster in the extracellular half of the protein similar to that in bR.

Significant differences to bR are found for the residues which are involved in its proton uptake and release. In the cytoplasmic half, Asp96 is replaced by Phe86 resulting in a significantly more hydrophobic environment, which explains the lower pumping activity of ppR compared to bR [387]. The proton release dyad (Glu194 and Glu204) in the extracellular side of bR is replaced by a single aspartate (Asp193).

The binding between ppR and its transducer protein (HtrII) is established via the helices F and G of ppR and the transmembrane helices TM1 and TM2 of the transducer. The helices G and TM2 are oriented parallel, while TM1 bends away from the receptor. The interface is formed mostly by van der Waals contacts and few hydrogen bonds, which are observed between Tyr199 (helix G) and Asn74 (TM2) as well as between Thr189 (FG loop) and Glu43 (TM1) and Ser62 (TM2) [376].

[‡]The nomenclature is adopted from refs [27, 325]: the first residue label and the residue number refer to ppR, and the second residue label specifies the corresponding amino acid of bR (ppR→ bR).

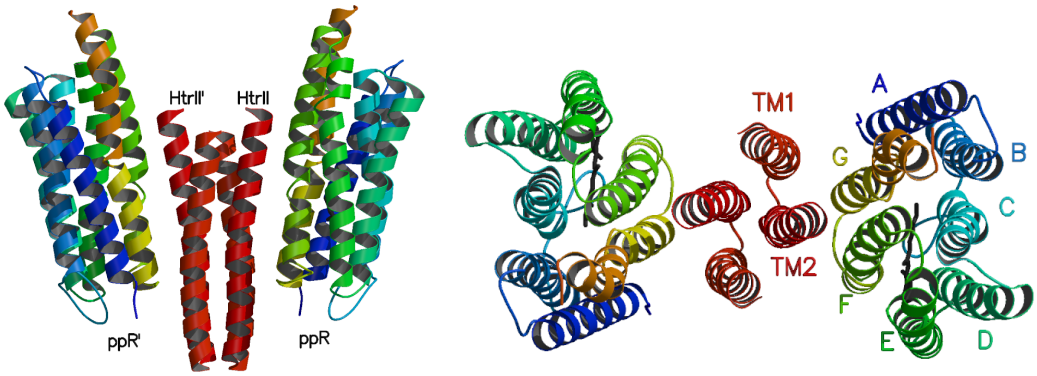


Figure 6.5: Structure of the ppR/HtrII complex consisting of two molecules of ppR and two molecules of HtrII (red) (PDB code: 1h2s). Left: Side view of the receptor/transducer complex. Right: Top view from the cytoplasmic side.

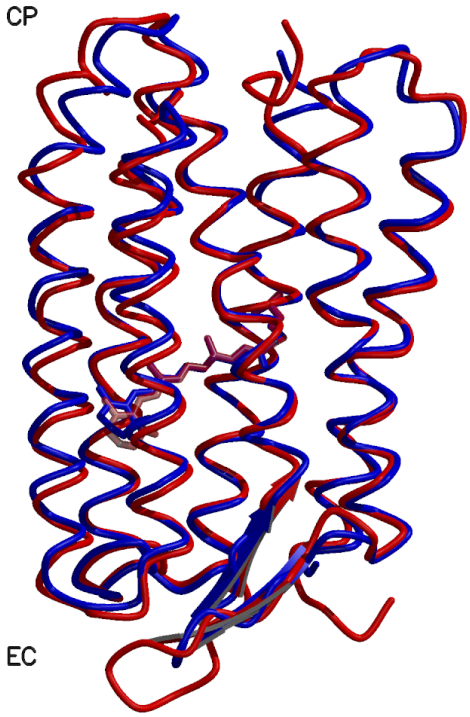


Figure 6.6: Superposition of the structure (C_α traces) and retinal of bR (red, 1c3w) and ppR (blue, 1h68). The antiparallel β-sheet at the BC loop is additionally shown in a strand representation in order to highlight differences.

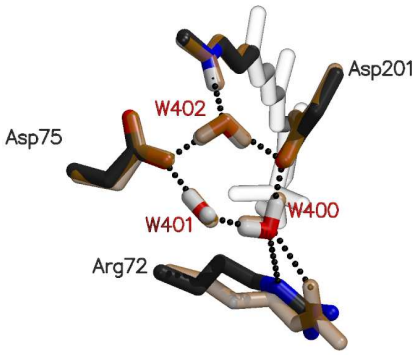


Figure 6.7: Detailed view of the retinal Schiff base region of ppR (PDB code: 1h68) superposed with that of bR (orange, 1c3w) showing the retinal and its counterion residues as well as the three water molecules of the hydrogen-bonded network.

Photocycle

Light activated all-*trans*/13-*cis*-isomerization of retinal triggers a photocycle [321, 372] which resembles mostly that of bR. The subscripts denote the wavelength of the visible absorption maximum λ_{max}.

$$ppR_{498} \longrightarrow K_{540} \longrightarrow L_{488} \longrightarrow M_{390} \longrightarrow O_{560} \longrightarrow ppR_{498}$$

The intermediate states show similarities and are named analogously to those of the bR photocycle. Additionally to the intermediates above, two different groups found a KL [321] or a N intermediate [20]. Furthermore, a slightly different photocycle consisting of eight kinetic states whereas some of them are an equilibrium mixture of intermediates was suggested [20]. Apart from these ambiguities, the time needed to complete the photocycle of ppR is significantly longer than that for bR. While the M intermediate is formed in the microsecond time scale as in bR, its decay back the ground state is orders of magnitude slower (second time scale [321]) than in bR.

The M state of ppR is characterized by an unprotonated Schiff base as in bR and the formation of the active signalling state (late M), which persists until the O intermediated decays [388–390]. The active signalling state is structurally characterized by displacements of the helices F and G caused by a loss of interhelical connections [375, 377, 389]. These displacements induce the signal transfer to the transducer via the interface of the helices F, G, and TM2, which eventually leads to a rotation of about 15° [375] and a tilt of TM2 [377].

As outlined above, ppR transports protons across the membrane if it is expressed without the transducer protein. The proton uptake then occurs during the M→O transition and the release during the O→ppR transition [321]. But even if ppR is associated with the transducer, light-induced proton uptake and release but no vectorial transport could be observed [381, 389], indicating a proton circulation (uptake and release from the same side).

6.3 Color Regulation - Current State of Knowledge

The mechanism of color tuning is based on the interaction of the chromophore with the electrostatic field of the protein environment and the mechanical strain applied on the chromophore by the environment. At present, the following mechanisms are proposed to contribute to the wavelength regulation in rhodopsins and the opsin shift in general:

(i) coplanarization of the ring-chain system and distortion of the chromophore structure[45, 47, 391–395],

(ii) electrostatic interaction of the chromophore with ionic, polar, and polarizable groups of the protein environment [21, 30, 31, 36, 37, 41, 43, 47–49, 396–399], and

(iii) a change of the interactions between the chromophore and its complex counterion [25, 37–39, 46, 400]

The degree of coplanarization of the β-ionone ring/polyene chain system determines the extend of the correlated π-system and hence leads to a bathochromic shift of the excitation energy relative to a twisted conformation of retinal. Among archaeal rhodopsins, this contribution is limited since retinal adopts a planar 6-s-*trans* ring/chain conformation in all archaeal rhodopsins. The contribution to the opsin shift, however, is significant [391] due to the twisted 6-s-*cis* conformation of retinal found in solution [44, 45]. Generally, this mechanism is connected with the conformational flexibility of retinal. In various rhodopsins highly twisted conformations are found.

The electrostatic interaction of the chromophore with ionic, polar, and polarizable groups of the protein environment destabilizes (stabilizes) the ground (excited) state of the chromophore leading to a hypsochromic (bathochromic) shift. This mechanism is connected to the high electronic polarizability of retinal and its large change in dipole moment upon excitation. The importance of the interaction with polar side chains is impressively demonstrated in human cone pigments [26, 48]: Only seven alcoholic groups (Thr, Ser) around the chromophore determine the shift between the red and green pigment. In addition to charged and polar groups, polarizable groups such as tryptophane contribute to the opsin shift because of their polarization associated with the large change in retinals dipole moment upon excitation [30, 53, 60, 401]. The influence on the spectral shift among archaeal rhodopsins, however, is supposed to be small [50], since the aromatic residues in proximity of the chromophore are conserved.

A change of the interaction between the PSB of retinal and its (complex) counterion may in principle be achieved by a variation of the distance between them, or indirectly by a change of the dielectric environment around the PSB.

Bacteriorhodopsin vs *pharaonis* Phoborhodopsin

The crystal structures of bR [332] and ppR [320, 378] show differences in the retinal geometry. However, FTIR spectroscopy [402] and previous theoretical calculations [50, 51] agree that mechanism (i) appears to be of minor importance in the case of bR vs ppR. Hence, the changes in chromophore geometry cannot explain the spectral shift between bR and ppR.

Mutation experiments have elucidated the role of mechanism (ii) in the color tuning by identifying residues in the retinal binding pocket which are involved in the spectral shift[27, 322–325, 403–406]. In particular, Shimono and co-workers [27, 322–325] have extensively studied the differences between bR and ppR. Single mutation experiments of residues in the binding pocket of ppR [27, 322] showed that each side chain has only small contributions to the color tuning. Even simultaneous mutation of multiple side chains could only produce parts of the spectral shift [27, 322]. A prominent example is the ppR mutant 'bR/ppR' with a binding pocket identical to that of bR, i.e., all ten different residues within 5 Å of the chromophore were replaced by the corresponding ones in bR. For this multiple mutant, about 40% of the shift was obtained.

Since mutation of even the whole binding pocket of ppR did not yield the complete spectral shift, chimeric proteins, in which either individual or multiple helices of one

protein were completely substituted by the corresponding helix of the other protein were investigated [324]. By substituting four of the seven helices, the total shift could approximately be reproduced and it was suggested that the interaction of helices D and E and the loop between them play a major role in color tuning.

These results indicate that mechanism (ii) is insufficient to explain the total spectral shift if only a single or a small number of residues are considered.

Mechanism (iii), the change in the interactions between the chromophore and its complex counterion or hydrogen-bond acceptor, has also found to be of importance. On the basis of frequency shifts of the Schiff base vibrations, Kandori and co-workers[323, 402, 407] found that the hydrogen bond between the protonated Schiff base and the counterion complex is stronger in ppR than in bR. Interestingly, FTIR measurements on the bR/ppR mutant indicated that the strength of the counterion interaction is the same as in ppR, although the binding pocket is identical to that of bR. This result shows that the mechanisms (ii) and (iii) are indeed working independently. Thus the bR/ppR mutant provides the opportunity to distinguish between the contributions of mechanisms (ii) and (iii). Furthermore, two of the counterion residues (the aspartic acids) and three water molecules form a pentagonal cluster structure, a hydrogen bonded network. This pentagonal structure was proposed to be distorted in ppR in contrast to bR, which was later confirmed by crystal structures [320, 332, 378]. So the question arises, whether and how these observations are related to the λ_{max} difference.

In a work by Shimono et al. [324], the distorted hydrogen bonded cluster was attributed to the different position of the guanidinium group of arginine 72 (Arg72) in ppR – it points toward the extracellular side in ppR forming a salt bridge with aspartate 193 (Asp193), while the corresponding residue in bR, Arg82, points toward the cytoplasmic side and thus interacts more strongly with the HBN. The importance of Arg72 was also pointed out by Luecke et al. [320] in their work about X-ray structures of ppR.

Ren et al.[50] came to a similar conclusion about the importance of Arg72 in their theoretical work based on the X-ray structures and quantum mechanical calculations for active site models. The different positions of the guanidinium group of Arg82/72 (bR/ppR) were suggested to be the main reason for the spectral shift because they influence the charge distribution on the counterion residues differently in bR and ppR. Recent experimental results [408, 409], however, have shown only small effects of Arg82/72 on the absorption maximum.

The combined QM/MM calculations of Hayashi et al. [51], in contrast, suggested that the spectral shift is induced by a shift of helix G that results in a shorter Schiff base–counterion residue (Asp201) distance in ppR, and hence a stronger chromophore–counterion interaction. However, this shorter distance is not apparent in the underlying crystal structure [378].

Kloppmann et al. [58] found significant differences in the electrostatic potentials at the retinal for ppR in comparison to bR and hR in agreement with the blue shifted absorption maximum of ppR. A decomposition of the electrostatic potential into contributions of individual residues suggested that seven residues are responsible for the differences in the electrostatic potentials.

So the theoretical studies came to different conclusions about the principal mecha-

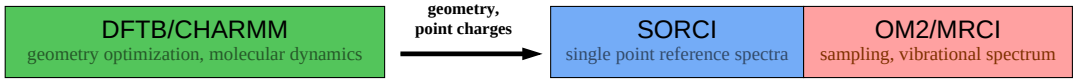


Figure 6.8: Computational strategy for the investigation of the color tuning in rhodopsins.

nism of spectral tuning – different positions of the guanidinium group of Arg82/72, a shorter Schiff base–counterion distance in ppR, or several residues located around the chromophore. The first two theoretical studies concluded that a major contribution to the spectral shift is due to geometrical differences of a particular amino acid in bR and ppR. This is important because it provides an explanation for the failure of mutational studies of the retinal binding pocket to completely reverse the spectral shift. However, the contributions from other amino acids of the binding pocket, which are experimentally known to account for approximately 40% of the shift [27], were not taken into account in these studies.

6.4 Computational Strategy and Methods

The general computational strategy employed to investigate the color tuning in rhodopsins is the result of an extensive assessment of various methodologies with regard to their performance in the description of the retinal chromophore, its ground-state geometry, optical properties, and response to electrostatic and steric interactions with its environment [410].

The overall approach is illustrated in Fig. 6.8 and based on ground-state geometry optimization and MD simulations in a combined QM/MM framework [136] using the approximate density functional method SCC-DFTB [138] for the QM region and the CHARMM27 force field [137] for the remainder of the protein. The calculation of excitation energies and excited state properties for the QM/MM structures is carried out using MRCI methods on the semiempirical (OM2/MRCI [147, 175, 411]) and ab initio (SORCI [412]) level of theory. The electrostatic potential of the MM region, as represented by fixed point charges of the CHARMM force field, is included in the corresponding QM Hamiltonians. The ab initio method SORCI is employed as highest computational level that is applicable to the full chromophore and few additional amino acid side chains. The semiempirical OM2/MRCI is employed for conformational sampling of the excitation energy.

In the following, the applicability of various methods to calculate excitation energies of retinal is shortly outlined. For a detailed review, see the joint publication with M. Wanko et al. [410]. The computational details regarding SCC-DFTB, OM2/MRCI, and SORCI will be introduced later in section 6.5.1.

The choice of the computational methods comprises three issues which have to be considered: the methods must be able to describe the geometrical properties of the chromophore properly; the methods must be capable to calculate the optical properties of the chromophore reliably, in particular with regard to conformational changes of the chromophore; and finally the methods must be able to reproduce the influence of the protein on the optical properties correctly.

The differences in the structure of the chromophore between different computational methods concern mainly the bond length alternation (BLA) of the polyene chain, calculated as the difference between the average single and double bond lengths, and the dihedral twist of the β-ionone ring. The BLA varies over a wide range for different methods [410]. While HF and post-HF methods, that mainly include static correlation, overestimate the BLA [410, 413], LDA/GGA-DFT methods, which primarily consider dynamic correlation, tend to underestimate the BLA. Hybrid functionals, however, in particular B3LYP and PBE0, are found to be in excellent agreement with CASPT2 [413, 414] which is regarded as best available method. The hybrid functionals thereby take advantage of an advantageous error cancellation between DFT (LDA) and HF.

SCC-DFTB has been shown to describe ground-state properties such as the BLA in good agreement with B3LYP, MP2, and CASPT2 calculations [410]. Therefore, it provides an alternative to the costly full ab initio methods. Additionally, SCC-DFTB yields ground-state torsional barriers for rotations around single and double bonds of retinal in agreement with B3LYP [415]. With regard to the dihedral twist of the β-ionone ring, HF and DFT methods, including SCC-DFTB, render almost planar structures for the all-*trans* retinal [410].

The accurate calculation of excitation energies of retinal proteins is challenging and various methods, such as time-dependent DFT (TD-DFT) [416], CASSCF [62, 417] and configurations interaction single (CIS), and multi-reference configuration interaction (MRCI) based on HF [56, 61, 62, 410, 416, 417] and semiempirical methods [50, 54, 55, 410] have been used.

Concerning the dependence of the excitation energy on geometrical parameters, pure DFT-based methods exhibit an erroneous dependence both on the BLA and on the dihedral twist of the β-ionone ring. In both cases, the errors are reduced if hybrid functionals are used. HF based methods (HF/CIS, SORCI) and semiempirical methods (e.g., OM2/MRCI), on the other hand, show the correct dependence [410]. Furthermore, investigations using models of retinal in rhodopsins, one for the 11-*cis* ground state and the other for the isomerized all-*trans* conformation (batho state) yielded that CASSCF and TD-DFT are not able to quantify the spectral shifts due to this geometrical change. While CASSCF overestimates and TD-DFT underestimates the spectral shift, CASPT2, SORCI, and OM2/MRCI are more appropriate [62, 410].

Similar results have been found for the dependence of the excitation energy on external field perturbations. TD-DFT fails dramatically to describe the effect of external charges. CASSCF and CIS methods (based on HF and semiempirical methods) are not reliable. While CIS methods underestimate the shift due to external charges, CASSCF overestimates the influence. On the other hand, CASPT2 and SORCI agree well and also the semiempirical OM2/MRCI method is able to predict spectral shifts due to changes in the electrostatic environment very accurately when compared to SORCI and CASPT2. Although absolute excitation energies are systematically overestimated.

Summarizing, the use of local exchange functionals in TDDFT as well as an unbalanced inclusion of dynamic and static correlation in methods like CASSCF or CIS are problematic to reliably predict spectral shifts for retinal. Instead, multireference approaches are required.

6.5 Color Tuning between Bacteriorhodopsin and Phoborhodopsin

In the following, the results of calculations on the spectral shift between bR and ppR are presented. After a short overview over the computational details (sec. 6.5.1), the spectral shift is explored based on QM/MM minimized structures in section 6.5.2 and 6.5.3. In section 6.5.4, a perturbation analysis is used to investigate the influence of each residue of the protein on the excitation energy. The impact of residues in the binding pocket is subsequently addressed in more detail in section 6.5.5 using mutation experiments. A vibrational analysis in section 6.5.6 and the calculation of the spectral shift on the basis of QM/MM dynamical simulations in section 6.5.7 finalize this study.

6.5.1 Computational Details

Ground-state geometry optimization and MD simulations were carried out in a combined QM/MM framework [136] using the approximate density functional method SCC-DFTB [138] for the QM region and the CHARMM27 force field [137] for the remainder of the protein. For the calculation of excitation energies and excited state properties for the QM/MM structures MRCI methods on the semiempirical (OM2/MRCI) and ab initio (SORCI) level of theory were used.

SCC-DFTB [138] has been applied in several QM/MM MD studies before [127, 130, 140–145]. With an efficiency comparable to semiempirical methods such as MNDO, AM1, or PM3, SCC-DFTB allows for long-time scale MD or MC simulations, which are not feasible at the ab initio or DFT level of theory.

The *Spectroscopy ORiented Configuration Interaction* (SORCI) method [412] is part of the ORCA quantum chemical package [418]. SORCI combines the concepts of classical multireference CI and multireference perturbation theory by dividing the first-order interacting space into weakly and strongly perturbing configurations. While the latter are treated variationally, the contributions of the former are included by second-order Møller-Plesset perturbation theory. The use of approximate natural orbitals eliminates the problem of choosing a suitable single-particle basis whose quality would affect the final CI result. SORCI gains computational efficiency by use of several thresholds which have been carefully adjusted for applications to the systems under study [†]. The employed basis set is Ahlrichs SV(P) [419] which is appropriate for calculations on the full chromophore.

The semiempirical OM2 Hamiltonian [147, 175] has been chosen for MRCI calculations [411] because for extended CI treatments this method is expected to be superior to traditional semiempirical Hamiltonians like MNDO or AM1 which underestimate the HOMO-LUMO gap due to their neglect of nonorthogonality between the atomic basis functions [147]. By applying orthogonalization corrections to the Fock matrix, OM2 overcomes this problem and thus yields improved MRCI excitation energies (as has been shown in ref [420] for the case of butadiene) without any reparametrization. The OM2/MRCISD calculations in the following were performed with version

[†]The thresholds used in this work are as follows: $T_{\mathrm{Pre}} = 10^{-3}$, $T_{\mathrm{Nat}} = 10^{-6}$, and $T_{\mathrm{Sel}} = 10^{-6}\mathrm{E_h}$ (see Supporting Information of ref [410]). Only core orbitals were frozen.

6.1 of the MNDO99 program using an active orbital window of 15 occupied and 15 virtual orbitals that are selected from a preliminary CI run. No individual configuration selection was applied.

The coordinates of bR and ppR were obtained from the X-ray crystallographic structures of Luecke et al. [332] (PDB code 1C3W) and Royant et al. [378] (PDB code 1H68), respectively. For bR, standard protonation states were assumed except for Asp96, Asp115, and Glu204, which were modeled in their protonated form [360, 421]. In ppR, all titratable residues were assumed to be charged [51]. Hydrogen atoms were added using the HBUILD module of the CHARMM program package [139]. The systems contain 3694 atoms and 23 crystal water for bR and 3495 atoms and 27 crystal water for ppR.

During geometry optimizations and molecular dynamics (MD) simulations, C_α-atoms with a distance to the Schiff base greater than 14 Å were harmonically restrained to their initial positions. All atoms were optimized or propagated during geometry optimizations and MD simulations.

To mimic the screening effect of bulk solvent, a Poisson-Boltzmann (PB) charge scaling scheme was used as proposed by Dinner et al. [422]. The method scales down the partial charges for charged residues exposed to the solvent on the surface of the protein according to a set of PB calculations [229].

The QM segment comprises retinal (63 atoms, charge +1), the boundary between the QM and MM region was chosen at the C_α-C_β bond and link atoms were introduced to saturate the valence of the QM boundary atoms. The interactions at the boundary are treated using the divided frontier charge scheme recently proposed and tested in a joint publication with P. H. König et al. [212]. For calculations with QM regions of different size, the QM segment comprises retinal for model A, retinal and the side chains Asp85/75 and Asp212/201 as well as the water molecules W402/402, W401/401 and W406/400 for model B (86 atoms, charge -1) and additionally Arg82/72 for model C (105 atoms, charge 0). For the electrostatic description of the MM part, electrostatic force shifting with a cutoff of 13.0 Å was used for geometry optimizations and molecular dynamics simulations.

The QM/MM dynamics simulations employed constant-temperature MD defined by the Nose-Hoover equations [423, 424] of motion, with a time step of 1.0 fs and a temperature of 300 K. In total 1.8 ns, split into three independent simulations that started from different random seeds, were performed for both proteins. For each simulation an equilibration run of 100 ps was followed by a 500 ps production run.

On the basis of the trajectories, the absorption spectra were obtained by calculating the excitation energies with OM2/MRCI along the trajectories [55] and using the corresponding histograms as an approximation to the absorption spectra. For each histogram 1000 excitation energies with a time interval of 0.1 ps were collected and a bin width of 0.05 eV was used. The structures were obtained from the last 100 ps of one of the simulations for bR and ppR, respectively. The vibrational properties were calculated using the autocorrelation function of the velocity obtained from QM/MM dynamics [425].

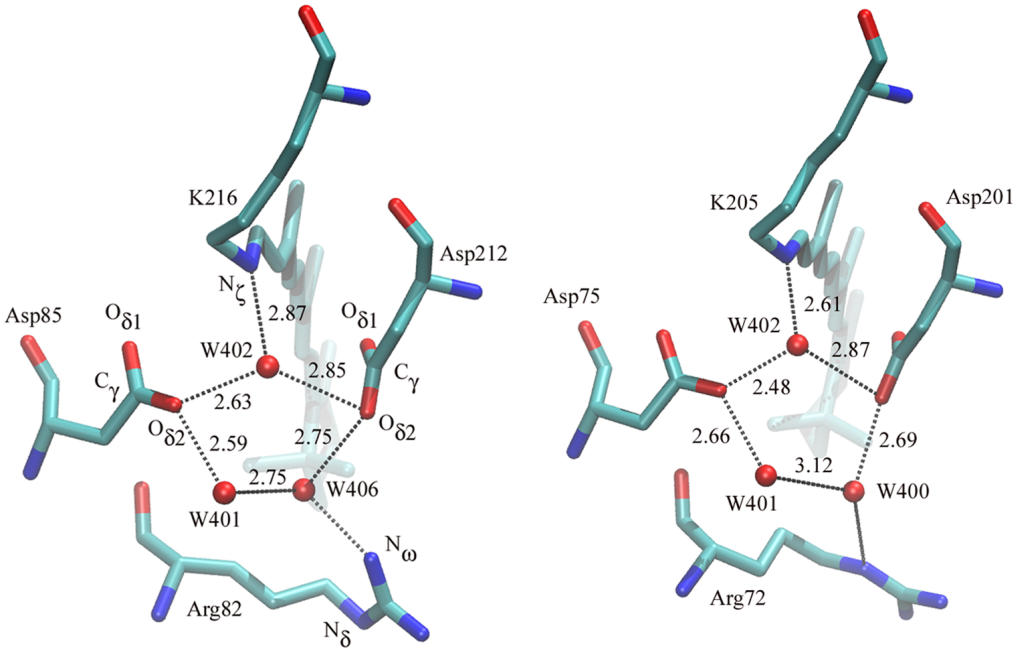


Figure 6.9: Hydrogen bonded network of bR (1C3W, left) and ppR (1H68). (Distances in Å)

6.5.2 Comparison of Structures

Geometry optimization of the X-ray structures of bR and ppR (PDB codes: 1C3W and 1H68) using the SCC-DFTB/CHARMM scheme yields nearly identical geometries for the chromophore. Previous theoretical studies [50, 51] came to similar results, while the crystal structures [320, 330, 331] show larger deviations between bR and ppR in the lysine side chain to which the retinal is covalently bound (see Hayashi et al. [51] for more details). Despite the similarity of the chromophore structures, the bond length alternation (BLA) of retinal, calculated as the difference between the average single and double bond lengths along the polyene chain, is clearly different for the QM/MM optimized structures of bR (0.056 Å) and ppR (0.065 Å). This value has a direct relationship to the excitation energy, as discussed in detail later.

Further, the distance between the β-ionone ring and the C_ζ atom of Arg82/72 is different in bR and ppR due to the different orientation of the guanidinium group in bR and ppR (Fig. 6.9). A detailed structural comparison of the complex consisting of the counterion residues and the three water molecules in the retinal binding pocket (which together are referred to as extended HBN in the following text) is summarized in Table 6.1.

Experimental studies [426] emphasized a structural difference in the HBN for bR and ppR (Fig. 6.9). Comparing the X-ray structures of the HBN, the ppR (1H68) structure shows a 'distortion' (with respect to bR) as indicated by a larger W401-W406/W401-400 distance, a larger distance between C_ζ of Arg82/72 and the Schiff base, and a shorter distance of the Schiff base nitrogen to the oxygen of W402 in ppR. These differences are not as pronounced in the QM/MM optimized structures presented here.

Table 6.1: Selected geometrical parameters of the retinal binding site for QM/MM optimized structures of this study, X-ray structures, and from previous theoretical studies. (Distances in Å; the two residue numbers specified refer to bR and ppR, respectively)

		present work		X-ray structures					previous works	
		bR	ppR	1C3W (bR)	1BRX (bR)	1QHJ (bR)	1H68 (ppR)	1JGJ (ppR)	Vreven[a] (bR)	Hayashi[b] (bR)
RET_{N_ζ}	O(W402/402)	2.81	2.82	2.87	2.68	2.85	2.61	2.64	2.80	2.63
O(W402/402)	$O_{\delta 2}$(Asp85/75)	2.57	2.59	2.63	2.80	2.57	2.48	2.40	2.55	2.58
O(W402/402)	$O_{\delta 2}$(Asp212/201)	2.75	2.71	2.85	3.36	3.17	2.87	3.06	4.11	4.07
RET_{N_ζ}	$O_{\delta 1}$(Asp85/75)	3.66	3.67	3.79	4.29	3.65	4.07	4.15	3.59	–
RET_{N_ζ}	$O_{\delta 2}$(Asp85/75)	4.33	4.36	4.38	4.19	4.17	4.28	3.98	3.91	–
RET_{N_ζ}	C_γ(Asp85/75)	4.35	4.41	4.45	4.66	4.28	4.58	4.42	–	–
RET_{N_ζ}	$O_{\delta 1}$(Asp212/201)	4.02	4.28	3.74	3.66	3.96	4.05	3.82	3.77	–
RET_{N_ζ}	$O_{\delta 2}$(Asp212/201)	4.77	4.71	4.85	4.87	5.18	4.63	4.89	–	–
RET_{N_ζ}	C_γ(Asp212/201)	4.29	4.36	4.12	3.99	4.41	4.11	4.25	–	5.2[c]
O(W402/402)	$O_{\delta 1}$(Asp85/75)	3.56	3.55	3.65	4.20	3.71	3.67	3.93	–	–
O(W402/402)	$O_{\delta 1}$(Asp212/201)	3.48	3.71	3.26	3.55	3.32	3.65	3.35	2.59	2.82
O(W401/401)	$O_{\delta 2}$(Asp85/75)	2.60	2.59	2.59	2.57	2.81	2.66	2.72	2.62	2.63
O(W401/401)	O(W406/400)	2.74	2.73	2.75	–	3.01	3.13	3.19	2.67	2.77
O(W406/400)	$O_{\delta 2}$(Asp212/212)	2.61	2.65	2.75	–	2.63	2.69	2.50	2.58	2.78
O(W406/400)	$N_{\omega/\delta}$(Arg82/72)	2.77	2.89	2.49	–	2.58	3.03	3.74	–	2.72
RET_{N_ζ}	C_ζ(Arg82/72)	10.32	10.15	9.39	8.46	9.71	10.14	10.45	–	–
RET_{C6}	C_ζ(Arg82/72)	11.59	10.91	11.6	10.99	11.60	10.92	11.10	–	–

[a] Ref [416]. [b] Refs [51, 417]. [c] Value taken from ref [51].

The deviation of the presented structures from the underlying crystal structures is within the margin of fluctuations between the available experimental structures (see Table 6.1).

Theoretical studies by Hayashi et al. [417] and Vreven et al. [416] yielded QM/MM optimized structures for bR that differ from the QM/MM optimized structure of this work and the crystal structures regarding the HBN in one respect [‡]: the water molecule W402 is hydrogen bonded to the O_{δ_1} atom of Asp212, rather than the O_{δ_2} atom (see Fig. 6.9 for notation) as in the QM/MM optimized structure of this work and the crystal structures (1BRX, 1C3W) (Table 6.1). This deviation from the crystal structures was already pointed out in ref [417]. As a consequence, the distances between the oxygen of W402 to O_{δ_1} and O_{δ_2} differ significantly (e.g., 2.82 and 4.07 Å in ref [417]), while this difference is much smaller in the presented QM/MM optimized structures (Table 6.1). Moreover, the distance between the Schiff base nitrogen and C_γ of Asp212 is increased from 4.12 Å (1C3W) to 5.2 Å in Hayashi's structure [51]. This was proposed to be responsible for the difference in excitation energy between ppR and bR, since the corresponding distance in ppR is clearly shorter (4.5 Å). In the QM/MM optimized structures of bR and ppR of this work, this distance is similar in both cases (4.29 Å in bR, 4.36 Å in ppR), in good agreement with the values found in the various X-ray structures.

In the theoretical study by Ren et al. [50], the geometry of the chromophore and the positions of the hydrogens were optimized while the positions of all other atoms were held fixed during the optimization. For the optimized chromophore a similar structure was found in bR and ppR. Other geometrical parameters like the structure of the HBN (and, e.g., the orientation of the guanidinium group of Arg72 in ppR) are the same as in the underlying crystal structures because of the chosen constraints.

All theoretical methods predict that the chromophore geometries in bR and ppR are very similar, except for the extent of BLA. The geometry of the HBN differs in the X-ray, the SCC-DFTB/CHARMM, and the other theoretical structures, and the effect of these geometrical differences on excitation energies will be examined in the next section.

6.5.3 Excitation Energies and Dipole Moments of bR and ppR

Excitation energies were calculated using SORCI and OM2/MRCI on the basis of the QM/MM optimized structures (Table 6.2). In the following, the chromophore is treated quantum mechanically while the rest of the protein is represented by point charges as taken from the CHARMM force field.

The spectral shift between bR and ppR is 0.29 eV with SORCI and 0.30 eV with OM2/MRCI (Table 6.2), both values being very close to the experimental one of 0.32 eV [20, 427]. The absolute energies at the SORCI level (2.34 eV for bR, 2.63 eV for ppR) are overestimated by about 0.1 eV. OM2/MRCI overestimates the excitation energies (2.66 eV for bR, 2.96 eV for ppR) by ∼0.3 eV compared to SORCI

[‡]Other geometrical variations are in the order of 0.3 Å and of minor importance. They may be due to differences in the applied quantum mechanical potentials or to local minima found by different optimization techniques. As will be discussed below (sec. 6.5.7), these variables show fluctuations of similar magnitude along MD trajectories.

Table 6.2: Vertical excitation energies $\Delta E_{S_1-S_0}$ (in eV).[a]

Method	System	bR	ppR	ΔE_{ppR-bR}
exp [20, 427]	protein	2.18	2.50	0.32
SORCI	vacuum[b,c]	1.86	1.91	0.05
	protein[b,d]	2.34	2.63	0.29
	$\Delta_{protein-vacuum}$	0.48	0.72	–
OM2/MRCI	vacuum[b,c]	2.22	2.24	0.02
	protein[b,d]	2.66	2.96	0.30
	$\Delta_{protein-vacuum}$	0.44	0.72	–
	extended HBN[b,e]	2.98	3.05	0.07
	$\Delta_{HBN-vacuum}$	0.75	0.81	–
	constrained protein[d,f]	2.66	2.99	0.33
	constrained HBN[e,f]	2.96	3.09	0.13

[a] From QM calculations on the chromophore. [b] Geometries from QM/MM optimization of the protein. [c] External point charges not included in QM calculations. [d] External point charges included in QM calculations. [e] Only point charges of Asp85/75, Asp212/201, Arg82/72, W402, W401 and W406/400 included in QM calculation. [f] Constrained geometry optimization: the chromophore was allowed to relax while the Schiff base nitrogen and the rest of the protein were fixed at the X-ray positions.

(Table 6.2), but reproduces relative energies in good agreement with SORCI and the experimental results as expected from our previous study [410].

Generally, the applied QM/MM treatment involves some approximations, in particular the neglect of environmental polarization and dispersion interactions that would redshift the excitation energy. These contributions have been addressed only by few theoretical approaches [50, 53, 55] and they have discussed in some detail previously [410]. For bR and ppR, Ren et al. [50] reported dispersive red shifts of 0.15 and 0.12 eV, resulting in a very small contribution of 0.03 eV to the spectral shift.

The dipole moment may be used to assess the global electrostatic interaction between the chromophore and the protein environment. As known from experiment, the dipole moment of the chromophore changes significantly upon excitation [428]. A large change in dipole moment $|\mu_{S_1} - \mu_{S_0}|$ is indeed computed both by SORCI (10.0 D for bR, 7.9 D for ppR) and by OM2/MRCI (11.8 D for bR, 9.4 D for ppR) (Table 6.3). These values are in qualitative agreement with those reported by Hayashi et al. [51] (15.0 D for bR, 10.1 D for ppR), while in the work of Ren et al. [50] the change is smaller in bR (8.4 D) than in ppR (12.4 D) However, these data are difficult to compare since in our approach and in Hayashi et al. [51] the dipole moment is calculated for the charged chromophore, while Ren et al. [50] reported the dipole moments of the complete, neutral binding pockets. Compared to OM2/MRCI and SORCI, CASSCF calculations [51] predict larger changes of the dipole moment upon excitation due to the lack of dynamical correlations [410]. For the same reason the influence of external charges on the excitation energy is overestimated with CASSCF [410].

The dipole moment of the chromophore is computed to be significantly different in the protein and in vacuum (i.e., with and without MM point charges). The cor-

Table 6.3: Dipole moments μ of the chromophore (in Debye) for QM/MM minimized structures [a].

	bR				ppR			
	μ_x	μ_y	μ_z	$\lvert\mu\rvert$	μ_x	μ_y	μ_z	$\lvert\mu\rvert$
SORCI								
$\mu_{S_0}^{\text{protein}}$ [b,d]	3.0	12.2	-5.7	13.8	3.3	8.3	-3.5	9.6
$\mu_{S_1}^{\text{protein}}$ [b,d]	3.8	20.2	-11.6	23.6	4.4	15.0	-7.6	17.4
$\mu_{S_0}^{\text{vacuum}}$ [b,c]	4.5	18.7	-6.6	20.3	5.3	18.0	-5.0	19.4
$\mu_{S_1}^{\text{vacuum}}$ [b,c]	5.0	26.2	-12.9	29.6	6.2	27.3	-11.2	30.1
$\mu_{S_1}^{\text{protein}} - \mu_{S_0}^{\text{protein}}$	–	–	–	10.0	–	–	–	7.9
$\mu_{S_0}^{\text{protein}} - \mu_{S_0}^{\text{vacuum}}$	–	–	–	6.8	–	–	–	10.1
$\mu_{S_1}^{\text{protein}} - \mu_{S_1}^{\text{vacuum}}$	–	–	–	6.3	–	–	–	12.9
OM2/MRCI								
$\mu_{S_0}^{\text{protein}}$ [b,d]	1.2	11.1	-4.8	12.2	1.9	8.2	-2.6	8.8
$\mu_{S_1}^{\text{protein}}$ [b,d]	2.0	20.4	-12.1	23.8	2.9	15.9	-7.9	18.0
$\mu_{S_0}^{\text{vacuum}}$ [b,c]	3.2	19.3	-6.0	20.5	4.4	19.9	-4.4	20.9
$\mu_{S_1}^{\text{vacuum}}$ [b,c]	3.5	27.3	-13.3	30.6	5.1	29.6	-11.5	32.2
$\mu_{S_1}^{\text{protein}} - \mu_{S_0}^{\text{protein}}$	–	–	–	11.8	–	–	–	9.4
$\mu_{S_0}^{\text{protein}} - \mu_{S_0}^{\text{vacuum}}$	–	–	–	8.6	–	–	–	12.2
$\mu_{S_1}^{\text{protein}} - \mu_{S_1}^{\text{vacuum}}$	–	–	–	7.1	–	–	–	14.4

[a] From QM calculations on the chromophore. [b] Geometries from QM/MM optimization of the protein. [c] External point charges not included in QM calculation. [d] External point charges not included in QM calculation.

responding differences $|\mu^{\text{protein}} - \mu^{\text{vacuum}}|$ for the S_0 ground state of bR and ppR are 6.8 and 10.1 D with SORCI, and 8.6 and 12.2 D with OM2/MRCI, respectively (the values for S_1 are similar). This shows that the ppR environment polarizes the chromophore more strongly than the bR environment.

A general discussion of the electronic structure of retinal and retinal models has already been given by various authors (see e.g., refs [410, 429, 430]) and the calculations in this work do not yield any new insights in this regard. Therefore, it is referred to the Supporting Information of ref [431] for a Mulliken population analysis. Note that the state "S_1" labels the bright, singly excited state throughout this work, although in ppR, the dark, doubly excited state is predicted by the SORCI method to be slightly lower in energy. A similar sequence of the two states in the presence of a strong counterion has been found in CASPT2 studies of a model chromophore in rhodopsin [432]. Note that in contrast to the CASSCF/CASPT2 results in ref [57], any mixing of the diabatic components between S_1 and S_2 or any transfer of oscillator strength to the doubly excited state is not observed [§].

[§] Whether the predicted state crossing is realistic or not, cannot be safely concluded from the current calculations, since the accuracy of SORCI and CASPT2 for the doubly excited state energy may be

In the remainder of this section, different factors controlling the spectral characteristics are investigated.

Role of the Chromophore Geometry. To elucidate the role of the chromophore geometry, excitation energies were calculated for the bare chromophore using either a gas-phase optimized structure or the QM/MM optimized structure of the chromophore in bR and ppR. Additionally, the nonoptimized crystal structures were used to check the effect of the minimization on the excitation energy of the bare chromophore.

For the gas-phase optimized structure of the chromophore, we found excitation energies of 1.86 eV for SORCI and 2.16 eV for OM2/MRCI, which are similar to the values for the QM/MM optimized structures (Table 6.2). The latter ones deviate by less than 0.05 eV for SORCI and 0.08 eV for OM2/MRCI from those for the gas-phase optimized structures. This means that the strain exerted on the chromophore by its protein environment does not cause distortions that tune the excitation energy since the variation in the BLA is induced electrostatically and not sterically. Hence, neither SORCI (0.05 eV) nor OM2/MRCI (0.02 eV) find a significant difference between the excitation energies of the two QM/MM optimized chromophores. Further, also the dipole moments in vacuo ($\mu_{S_{0/1}}^{\text{vacuum}}$) of the chromophores (see Table 6.3) are nearly the same as expected because of their similar geometry.

For the nonoptimized chromophores taken from the crystal structures, the excitation energies (OM2: 2.23 eV for bR, 2.21 eV for ppR) agree well with those found for the QM/MM optimized structures, implying that the effect of geometry minimization is very small.

Influence of the Counterion Complex. The contribution of the counterion complex (Asp85/75, Asp212/201, and Arg82/72) to the electrostatic environment has been regarded as a key to understanding the mechanism of color tuning [50, 51, 402, 407]. To estimate the effect of the counterion complex on the spectral shift, the excitation energies were calculated using OM2/MRCI for a model only including the point charges of the extended HBN, i.e., of the counterion complex and the three water molecules in the retinal binding pocket (extended HBN in Table 6.2). For these extended HBN models, excitation energies of 2.98 eV for bR and 3.05 for ppR are found (Table 6.2). Hence, the difference of the excitation energies decreases to 0.07 eV, i.e., the contribution of the extended HBN to the total shift is only 23 %. The counterions obviously blue-shift the excitation energies strongly while the remainder of the protein causes a red-shift, which is larger in bR than in ppR.

As discussed above, the HBN is less distorted in the QM/MM optimized structures than in the crystal structures. To investigate the influence of the distortion, the excitation energies were recalculated using the crystal structures optimized under constraints: only the chromophore geometry was allowed to relax, while the position of the Schiff base nitrogen and the rest of the protein were fixed, thus preserving the distorted structure of the HBN. The excitation energies were computed incorporating either the whole protein or only the extended HBN as point charges (last two lines in Table 6.2).

For the whole protein, the spectral shift between bR and ppR increases by 0.03 eV

lower than for the singly excited one and no experimental evidence has been put forward about the first. Hence, further discussion of the doubly excited state are omitted.

to 0.33 eV, indicating that the slight differences in the counterion residue geometry and the distorted HBN have only little impact on the spectral shift. The difference in excitation energies for the extended HBN models increases from 0.07 eV as evaluated with the QM/MM optimized structures to 0.13 eV (Table 6.2). This shows that the distortion of the HBN structure has some influence on the excitation energy, which may be underestimated in our QM/MM model. Nonetheless, it confirms that the overall contribution of the counterion residues and the HBN to the spectral shift (about 30-40 %, or 0.1 eV) is much smaller than previously assumed.

Two preceding theoretical studies [50, 51] proposed structural differences in the (extended) HBN to cause the complete spectral shift. Hayashi et al. [51] attributed the spectral shift to an increased distance between the Schiff base nitrogen and Asp212 in bR relative to ppR (5.2 Å in bR, 4.5 Å in ppR), whereas Ren et al. [50] hold the orientation of Arg82/72 responsible for the effect. In contrast, the calculations in this work reproduce a major part of the spectral shift only after incorporating the remainder of the protein environment, which red-shifts the excitation energy in bR to a larger extent than in ppR.

In order to elucidate the differences between the models and the interpretation of the results further, it is helpful to analyze the electrostatic interactions between the chromophore and various parts of the protein environment.

The energy of the QM/MM system for a given electronic state and geometry contains three contributions that may change upon excitation (see also ref [51]): (1) the energy E^0 of the isolated chromophore, (2) the electrostatic (Coulomb) interaction energy $V^{\mathrm{ES}}_{\text{ch-pr}}$ between the polarized chromophore density and the protein, and (3) the electronic reorganization energy E^{ERO} which is needed to polarize the chromophore and is positive because of the variational principle. The interaction energy $V^{\mathrm{ES}}_{\text{ch-pr}}$ can be further divided into the interaction of the protein with the unpolarized chromophore ($V^{\mathrm{ES,vac}}_{\text{ch-pr}}$) and the change in $V^{\mathrm{ES}}_{\text{ch-pr}}$ due to the chromophore polarization ($V^{\mathrm{ES,pol}}_{\text{ch-pr}}$). The total energy gain due to polarization of the chromophore is the sum of $V^{\mathrm{ES,pol}}_{\text{ch-pr}}$ and E^{ERO}.

According to calculations in this work, both terms ($V^{\mathrm{ES}}_{\text{ch-pr}}$, E^{ERO}) contribute to the shift of ppR vs bR to the same degree (0.13 eV for $\Delta_{\text{ppR-bR}}E^{\mathrm{ERO}}$; 0.12 eV for $\Delta_{\text{ppR-bR}}V^{\mathrm{ES}}_{\text{ch-pr}}$). Thus, the contributions of individual residues to the excitation energy is investigated.

The best way to obtain these is by perturbational analysis which requires the calculation of the excitation energy by switching off the partial charges of every residue successively. This will be presented in section 6.5.4. As a faster alternative, the comparison of $V^{\mathrm{ES}}_{\text{ch-pr}}$ and $V^{\mathrm{ES,vac}}_{\text{ch-pr}}$ can be used to estimate the change in E^{ERO} that is caused by the perturbation from the charges of the individual amino acids. Under the assumption that $V^{\mathrm{ES,pol}}$ is invested in the electronic reorganization ¶ , $E^{\mathrm{ERO}} = -\frac{1}{2}V^{\mathrm{ES,pol}}_{\text{ch-pr}}$, where $V^{\mathrm{ES,pol}}_{\text{ch-pr}} = V^{\mathrm{ES}}_{\text{ch-pr}} - V^{\mathrm{ES,vac}}_{\text{ch-pr}}$ is obtained. Hence, only the ground-state and excited-state charges from two calculations, one including and one omitting the en-

¶This approximation, common in atomic polarization models, describes a medium which can be separated into linear polarizable centers [433].

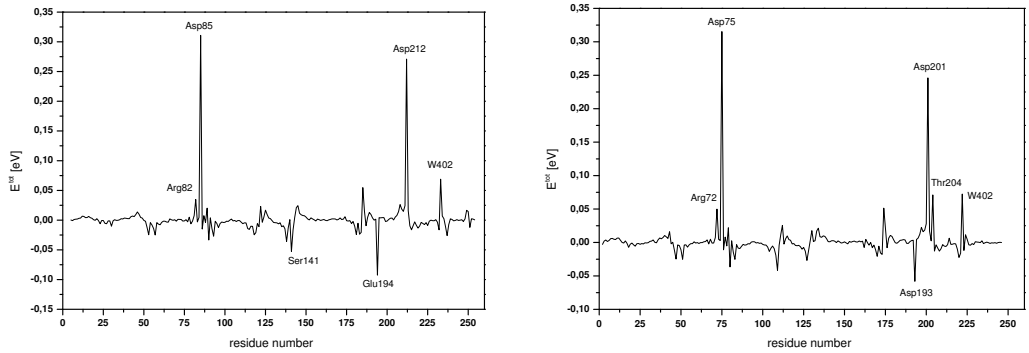


Figure 6.10: Calculated contributions to the excitation energy $\Delta_{S_1-S_0}E$ for individual residues for bR (left) and ppR (right) based on eq 6.2.

tire protein charges, are needed to obtain the contributions of individual residues to the total or excitation energy. $\Delta_{S_1-S_0}$ denotes the difference of a contribution, e.g., the electrostatic interaction energy $V^{\mathrm{ES}}_{\text{ch-pr}}$, between the S_1 and the S_0 state:

$$\Delta_{S_1-S_0}E = \Delta_{S_1-S_0}E^0 + \Delta_{S_1-S_0}V^{\mathrm{ES}}_{\text{ch-pr}} + \Delta_{S_1-S_0}E^{\mathrm{ERO}} \tag{6.1}$$

$$\approx \Delta_{S_1-S_0}E^0 + \tfrac{1}{2}(\Delta_{S_1-S_0}V^{\mathrm{ES}}_{\text{ch-pr}} + \Delta_{S_1-S_0}V^{\mathrm{ES,vac}}_{\text{ch-pr}}) \tag{6.2}$$

The results for bR and ppR are shown in Fig. 6.10. The largest contributions to the shift come from the difference between Glu194/Asp193, Ser141/Gly130, and Ala215/Thr204. Although the primary counterion residues Asp85/75 and Asp212/201 dominate the blue shift induced by the protein in both cases, their contribution to the shift of ppR vs bR is rather small, [‖] even negative (-0.025 eV) for the latter residue.

The results for the counterions residues imply that the counterion complex would not affect the ppR/bR shift at all which is in clear contrast to the results obtained with the HBN model and points out the limitations of an analysis based on eq 6.2.

Interestingly, qualitatively the same results are obtained using Hayashi's QM/MM optimized structures of bR and ppR [51]. In particular, the counterion complex contributes only very slightly (0.02 eV) to the shift.

In order to investigate the influence of the differences in the geometry of the HBN in the model of Hayashi et al. [51], as outlined in section 6.5.2, further, the excitation energies were recalculated using Hayashi's QM/MM optimized structures. In an analogous manner as for the analysis of the crystal structures, only the chromophore geometry was allowed to relax while the position of the Schiff base nitrogen and the rest of the protein was fixed, thus preserving the differences of the HBN.

[‖] The decisive role that Hayashi et al. [51] attribute to the counterion residue Asp212 results from an analysis of the contributions of the individual amino acids to the interaction energy between the ppR protein charges and differential chromophore charges $[q_{\mathrm{ch}}(\mathrm{ppR}) - q_{\mathrm{ch}}(\mathrm{bR})]$ for the S_1 state in which the chromophore is strongly polarized. The authors take this interaction energy as an estimate for the electrostatic contribution to the polarization energy, $V^{\mathrm{ES,pol}}_{\text{ch-pr}}$, which would be exact for $q_{\mathrm{ch}}(\mathrm{bR}) = q_{\mathrm{ch}}(\mathrm{vacuo})$.

Table 6.4: Vertical excitation energies $\Delta E_{S_1-S_0}$ (in eV) for previous theoretical structures [51] using the OM2/MRCI method [a].

	Hayashi		
	bR	ppR	$\Delta E_{ppR\text{-}bR}$
protein[b,c]	2.65	3.00	0.35
HBN model[b,d]	2.91	3.01	0.10

[a] From QM calculations on the chromophore. [b] Geometries from constrained QM/MM optimization of the protein. [c] External point charges included in the QM calculation. [d] Only point charges of Asp85/75, Asp212/201, Arg82/85, W402, W401 and W406/400 included in QM calculation.

The calculated excitation energies and the spectral shift (Table 6.4) are similar to those evaluated for our QM/MM optimized and the X-ray structures for the whole protein as well as for the HBN model (Table 6.2). Hence, the increased distance between the Schiff base nitrogen and Asp212 in bR relative to ppR (5.2 Å in bR, 4.5 Å in ppR) in their model do not influence the spectral shift to the extent they concluded.

To investigate the influence of the different orientations of the counterion residue Arg82/72 (Fig. 6.9) on the spectral shift, as reported by Ren et al. [50], the calculations were repeated with QM regions of different size (model B and C), since the suggested influence relies on a stabilization of the chromophore in ppR by allowing Asp75 and Asp201 to redistribute their negative charge more toward the chromophore.

The different QM regions include retinal, both aspartates, and the water molecules of the HBN (model B), and finally the complete counterion complex plus the water molecules of the HBN (model C). The different QM regions were used for QM/MM geometry optimization of the whole protein. The geometrical parameters for the structures are given in Table 6.5. Excitation energies were calculated using OM2 for the bare QM region, i.e., without including the point charges of the remaining protein (Table 6.6).

The size of the QM region has only a small effect on the structures of the binding pocket; the distances in the hydrogen bonded network change moderately, reflecting cooperative effects in the HBN due to the quantum description. The excitation energies increase drastically from model A to model B (Table 6.6) as expected due to the incorporation of the negatively charged counterion residues Asp85/75 and Asp212/201. The spectral shift between bR and ppR is almost unaffected. Further increasing of the QM region to model C (adding Arg82/72) leads to a slight increase of the spectral shift from 0.03 eV for model B to 0.08 eV for model C.

The result for model C agrees with the results for the HBN models of the QM/MM optimized and X-ray structures (Table 6.2), which show that the orientation of the counterion residue Arg82/72 and more general the complex counterion is not solely responsible for the spectral shift. The results, therefore, confirm recent experimental results [408, 409] which also showed only a small effect of Arg82/72 on the absorption maximum.

Table 6.5: Selected geometrical parameters of the retinal binding site for QM/MM minimized structures of varying QM zones [a] (Distances in Å; the two residue numbers specified refer to bR and ppR, respectively).

		bR		ppR	
		B	C	B	C
RET_{N_ζ}	O(W402/402)	2.68	2.68	2.67	2.66
O(W402/402)	$O_{\delta 2}$(Asp85/75)	2.69	2.69	2.68	2.67
O(W402/402)	$O_{\delta 2}$(Asp212/201)	2.89	2.91	2.77	2.77
RET_{N_ζ}	$O_{\delta 1}$(Asp85/75)	3.44	3.47	3.58	3.65
RET_{N_ζ}	$O_{\delta 2}$(Asp85/75)	4.26	4.28	4.23	4.25
RET_{N_ζ}	C_γ(Asp85/75)	4.17	4.18	4.26	4.30
RET_{N_ζ}	$O_{\delta 1}$(Asp212/201)	3.77	3.76	4.12	4.12
RET_{N_ζ}	$O_{\delta 2}$(Asp212/201)	4.77	4.81	4.64	4.62
RET_{N_ζ}	C_γ(Asp212/201)	4.11	4.12	4.20	4.19
O(W402/402)	$O_{\delta 1}$(Asp85/75)	3.55	3.55	3.66	3.67
O(W402/402)	$O_{\delta 1}$(Asp212/201)	3.37	3.34	3.67	3.69
O(W401/401)	$O_{\delta 2}$(Asp85/75)	2.71	2.71	2.70	2.69
O(W401/401)	O(W406/400)	2.80	2.80	2.76	2.76
O(W406/400)	$O_{\delta 2}$(Asp212/201)	2.71	2.72	2.75	2.73
O(W406/400)	$N_{\omega/\delta}$(Arg82/72)	2.78	2.74	2.90	2.81
RET_{N_ζ}	C_ζ(Arg82/72)	10.03	9.92	9.87	9.86
RET_{C6}	C_ζ(Arg82/72)	11.59	11.46	10.90	10.76

[a] From QM calculations on retinal, Asp85/75, Asp212/201, water (model B); retinal, Asp85/75, Asp212/201, Arg82/72, water (model C).

Table 6.6: Vertical excitation energies $\Delta E_{S_1-S_0}$ (in eV) for varying QM zones [a] using the OM2 method.

model	bR	ppR	$\Delta E_{ppR\text{-}bR}$
A - vacuum[b]	2.22	2.24	0.02
B - vacuum[b]	2.70	2.73	0.03
C - vacuum[b]	3.03	3.11	0.08

[a] From QM calculations on the retinal (model A); retinal, Asp85/75, Asp212/201, water (model B); retinal, Asp85/75, Asp212/201, Arg82/72, water (model C). [b] External point charges not included in the QM calculation. Geometries from QM/MM optimization of the protein.

Table 6.7: Calculated OM2/MRCI shifts of the vertical excitation energies $\Delta\Delta E_{\text{mutant-wild type}}$ (in eV) and position for glycine-mutants of selected residues [a].

bR/ppR	bR	ppR	$\Delta E_{\text{ppR-bR}}$	position	bp
counterion residues					
Asp85/75	-0.39	-0.39	0.00	H-C	+
Asp212/201	-0.37	-0.32	0.05	H-G	+
Arg82/72	0.01	-0.02	-0.03	H-C	+
water molecules in the HBN[b]					
W401/W401	0.02	0.02	0.00	–	+
W402/W402	-0.09	-0.08	-0.01	–	+
W406/W400	0.03	0.02	0.01	–	+
remaining charged residues					
Glu194/Pro183	0.08	0.00	-0.08	H-F	–
GluH204/Asp193	0.01	0.04	0.03	H-G	–
conserved residues					
Trp182/171	-0.01(8)	-0.01(4)	0.00	H-F	+
Tyr185/174	-0.02(5)	-0.01(2)	0.01	H-F	+
Trp189/178	0.02	0.01	-0.01	H-F	+
Tyr57/51	0.02	0.02	0.00	H-B	–
Tyr83/73	0.01(8)	0.00(4)	-0.01	H-C	+
Trp86/76	0.03	0.03	0.00	H-C	+
Thr89/79	-0.03	-0.03	0.00	H-C	+
Thr90/80	0.04	0.03	-0.01	H-C	+

[a] The QM-region in OM2/MRCI consists of the chromophore. The position column indicates the location of the mutated residue, i.e., H-C indicates this residue to be located at helix C (etc.); the bp column specifies whether the residue is located in the binding pocket. [b] The partial charges of the water molecules were deleted for the calculation of the shift.

6.5.4 Perturbation Analysis

To analyze the influence of the individual amino acids, each residue was successively replaced by glycine and the excitation energy was recalculated using OM2/MRCI without further minimization of the protein. All amino acids having a notable influence (larger than 0.005 eV) on the excitation energy are discussed below. These amino acids can be grouped into four categories, (i) the counterion residues Asp85/75, Asp212/201, and the nearby Arg82/72, (ii) other charged residues located at the extracellular side of the protein, (iii) conserved residues in the binding pocket, and (iv) remaining neutral, nonconserved residues. The influence of the water molecules in the extended HBN was determined by simply setting the respective charges to zero. The results of this perturbation analysis are summarized in Table 6.7.

Influence of the Counterion Residues and the Extended HBN. The counterion residues Asp85/75 and Asp212/201 (in bR/ppR) cause large blue shifts of 0.3–0.4 eV. The effect of Arg82/72 is small and opposite in bR (-0.01 eV) and ppR (0.02 eV). The guanidinium group of Arg82/72 has a different orientation in the two proteins. It

is oriented toward the chromophore in bR, forming a hydrogen bond with the extended HBN, whereas it is pointing toward the extracellular side in ppR, forming a salt bridge with Asp193. The different orientation may cause Arg82/72 to affect the excitation energy in opposite direction. This is in qualitative agreement with the findings of Ren et al. [50]; however, since their calculations predict a much more pronounced effect, they proposed Arg82/72 to be responsible for the complete spectral shift.

The water molecules in the HBN seem to be of minor importance for the spectral difference between bR and ppR (Table 6.7).

The replacement of all three charged residues (Asp85/75, Asp212/201, and Arg82/72) plus the water molecules in their proximity causes a strong red shift of -0.66 eV for bR and -0.73 eV for ppR, i.e., omitting the HBN reduces the spectral shift between bR and ppR from 0.30 eV to 0.23 eV. The complete extended HBN contributes therefore 0.07 eV to the total spectral shift. This is consistent with the results in section 6.5.3 and confirms that the counterion residues account only for part of the spectral shift.

Note that the effects of the counterion residues and water molecules in Table 6.7 are not additive. Therefore, the sum of the individual effects differs from the results obtained through replacement of the complete extended HBN.

Influence of Other Charged Residues. The effect of all solvent exposed amino acids on the excitation energies is, in accordance with experimental results [341], small since their charges are reduced due to the solvent screening modeled by the charge scaling procedure.

Among the remaining charged residues, only the anions Glu194 in bR and Asp193 in ppR have a sizable influence (Table 6.7). Both residues red-shift the excitation energy, by 0.08 eV (Glu194 in bR) and 0.04 eV (Asp193 in ppR), respectively. In contrast to the counterion residues Asp212/201 and Asp85/75, they cause a red shift of the excitation energy because both are closer by 3-4 Å to the β-ionone ring than to the Schiff base nitrogen of the chromophore. They belong to a large HBN at the extracellular side of the protein that connects the Schiff base region with Glu194 and Glu204 in bR and with Asp193 in ppR (via several water molecules and Arg82/72).

The simultaneous replacement of the extended HBN and the charged groups Glu194/Asp193 leads to a red shift of -0.59 eV in bR and of -0.68 eV in ppR. Therefore, omitting the complex network of hydrogen bonded and charged residues at the extracellular side of bR and ppR reduces the spectral shift between bR and ppR by about 0.09 eV to a value of 0.21 eV. This is only slightly less than the value obtained for the extended HBN alone (0.23 eV, see above).

Influence of Conserved Residues in the Binding Pocket. The conserved aromatic tyrosines and tryptophanes have only a minor influence on the excitation energy and the spectral shift (Table 6.7). Similar results were found by Ren et al. [50] for Trp86/76, Trp182/171, and Tyr185/174. The two conserved threonines (Thr89/79, Thr90/80) do not influence the spectral shift either, as their effect is the same in bR and ppR.

Replacing all residues of this group simultaneously leads to blue shifts of 0.07 eV for bR and 0.05 eV for ppR, and hence to a differential shift of 0.02 eV. It should be noted again, however, that these calculations neglect the effects of protein polarization and

Table 6.8: Calculated OM2/MRCI shifts of the vertical excitation energies $\Delta\Delta E_{\text{mutant-wild type}}$ (in eV) and position for glycine-mutants of polar residues [a].

bR/ppR	bR	ppR	$\Delta E_{\text{ppR-bR}}$	position	bp	exp.
Trp138/Phe127	0.01	0.00	-0.01	H-E	+	+
Ser141/Gly130	0.07	–	-0.07	H-E	+	+
Thr142/Ala131	0.03	0.00	-0.03	H-E	+	+
Met145/Phe134	0.02	0.01	-0.01	H-E	+	+
Ile119/Met109	0.00	0.01	0.01	H-D	+	+
Ala215/Thr204	0.00	-0.07	-0.07	H-G	+	+
Met118/Val108	0.00(4)	-0.00(1)	-0.01	H-D	+	+
Thr121/Ala111	0.02	0.00	-0.02	H-D	–	-
Ser214/Val203	0.01	0.00	-0.01	H-G	–	-
Pro50/Ser44	0.00	0.01	0.01	H-B	–	-
Ser183/Ala172	0.01	0.00	-0.01	H-F	–	-
Thr55/Val49	0.00	0.00	0.00	H-B	–	-
Gly116/Thr106	–	-0.01	-0.01	H-D	–	-
all	0.16	-0.04	-0.20	–	–	–

[a] The QM-region in OM2/MRCI consists of the chromophore. The position column indicates the location of the mutated residue, i.e., H-C indicates this residue to be located at helix C (etc.); the bp column specifies whether the respective residue is located in the binding pocket.

dispersion, which were estimated [50] to cause an additional differential red shift of 0.03 eV (see section 6.5.3).

Influence of Polar Residues in the Binding Pocket. This group consists of all residues that have a sizable influence on the excitation energy and are neither charged nor conserved residues. Several of these residues (see last column in Table 6.8) have also been analyzed in experimental mutation studies [27, 325] which will be discussed in detail in section 6.5.5.

Among the residues of this group, only the replacements Ala215/Thr204, Ser141/Gly130, and Thr142/Ala131 lead to a substantial differential shift (Table 6.8). The replacements Thr121/Ala111, Pro50/Ser44, and Ser214/Val203 cause minor but noticeable differential shifts, even though they are further away from the chromophore and do not belong to the binding pocket. For the remaining substitutions the differential shift is less than 0.01 eV.

The simultaneous replacement of all polar residues (Table 6.8) leads to a blue shift of 0.16 eV in bR and a red shift of -0.04 eV in ppR. Thus these residues contribute 0.20 eV to the overall spectral shift of 0.30 eV. A summation of all individual contributions yields the same value (0.20 eV), indicating that cooperative effects play a minor role in this case.

6.5.5 Mutation Experiments in ppR

Several mutation studies, in particular those of Shimono and co-workers [27, 322–325], have addressed the role of various amino acids in the spectral tuning between bR and ppR. To compare the results of this work directly to these experiments, individual residues or groups of residues were mutated in ppR to their counterparts in bR. After geometry optimization of either the mutated residue(s) or the whole protein, the excitation energies were recalculated using OM2/MRCI (Table 6.9). Those of the most important mutations were also determined at the SORCI level. The SORCI calculations allow us to distinguish between the errors arising from the use of a more approximate QM method (OM2/MRCI) and from the applied QM/MM scheme, which neglects the polarization of the protein environment.

The results show (Table 6.9) that complete optimization of the protein after mutation yields larger shifts than optimization of the mutated residue(s) only. The agreement with the experimental shifts, even for the examined double, triple and multiple mutants, is fairly good using SORCI and the full relaxation scheme (ii), while the OM2/MRCI shifts are generally somewhat overestimated.

In the following, the discussion will focus on the OM2/MRCI values to facilitate comparisons with the preceding OM2/MRCI-based perturbation analysis (section 6.5.4).

Among the single mutations, Thr204Ala, Gly130Ser, Val108Met, Ala131Thr, and Ile43Val [**] lead to large effects. The largest impact is produced experimentally as well as theoretically by the mutations Thr204Ala (-0.06 eV; exp. -0.04 eV) and Gly130Ser (-0.06 eV; exp. -0.02 eV), whereas the other replacements (Val108Met, Ala131Thr, Ile43Val) are less effective, each with an experimentally measured shift of -0.015 eV (120 cm^{-1}).

In general, the absolute and relative contributions of these residues to the spectral shift have been predicted quite well by the perturbation analysis (Table 6.8) which gives significant contributions for the replacements Thr204Ala, Gly130Ser, and Ala131Thr (-0.07, -0.07, and -0.03 eV), and a minor one for Val108Met (-0.005 eV) that is underestimated but has the correct sign. The substitution Ile43Val does not contribute in the perturbation analysis because both residues are unpolar (and because the perturbation analysis does not include relaxation of the geometry).

The remaining five (Ile83Leu, Asn105AspH, Met109Ile, Phe127Trp, Phe134Met) mutations do not show any significant shift, neither in the experiment nor in our calculations, which is also in accord with the results from the perturbation analysis (sec. 6.5.4).

In addition to the single replacements, double, triple, and multiple mutants were also investigated. Of special importance is the triple mutant Val108Met/Gly130Ser/Thr204Ala, whose three sites are conserved among the long-wavelength rhodopsins hR, bR and sRI, but not in ppR, and are therefore expected to be involved in the spectral tuning of archaeal rhodopsins. However, the calculated shift (OM2/MRCI -0.14 eV, SORCI -0.11 eV) as well as the experimentally observed

[**]The nomenclature was adopted from refs [27, 325] and corresponds to the mutation ppR→ bR. The residue numbers refer to ppR.

Table 6.9: Calculated shifts of the vertical excitation energies $\Delta\Delta E_{\text{mutant-wild type}}$ (in eV) of various ppR mutants [a] calculated with OM2/MRCI and SORCI [b].

		OM2/MRCI		SORCI	
	exp.[27, 325]	local relaxation[c]	full relaxation	local relaxation[c]	full relaxation
Val108Met	-0.015	0.01	-0.02	0.00	-0.05
Gly130Ser	-0.02	-0.03	-0.06	-0.01	0.00
Thr204Ala	-0.044	-0.07	-0.06	-0.05	-0.05
Gly130Ser/Thr204Ala	-0.073	-0.11	-0.13	-0.07	-0.07
Val108Met/Gly130Ser	-0.049	-0.02	-0.07	-0.01	-0.06
Val108Met/Thr204Ala	-0.049	-0.06	-0.08	-0.04	-0.09
Val108Met/Gly130Ser/Thr204Ala	-0.077	-0.09	-0.14	-0.05	-0.11
Ile43Val	-0.015	0.00	-0.01		
Ile83Leu	-0.01	0.00	0.00		
Asn105AspH	0.00	-0.00	0.01		
Met109Ile	0.01	0.02	0.03		
Ala131Thr	-0.015	-0.01	-0.01		
Phe127Trp	0.005	-0.00	0.01		
Phe134Met	-0.005	-0.01	0.00		
bR/ppR	-0.118	-0.13	-0.19	-0.07	-0.16

[a] The nomenclature was adopted from refs [27, 325]: the first residue label and the residue number refer to ppR, and the second residue label specifies the corresponding amino acid of bR (ppR→ bR). [b] The QM-region consists of the chromophore. [c] Reminimization of the mutated residue only.

Table 6.10: Calculated shifts of the N-H stretch vibration and bond length alternation for bR, ppR and the bR/ppR mutant.

	$\nu(\mathrm{NH}) - \nu_{\mathrm{bR}}(\mathrm{NH})$ [cm^{-1}]	BLA [Å]	$\Delta E_{S_1-S_0}$ [eV]
bR	–	0.056	2.66
bR/ppR	-29	0.059	2.77
ppR	-25	0.065	2.96

shift (-0.08 eV) is only a fraction of the complete shift between bR and ppR (exp. $\sim$ 30 %).

A second important mutant is the multiple mutant bR/ppR, in which all ten different residues of the binding pocket within 5 Å are substituted such that the binding pocket of this mutant is identical to that of bR. Even so, the mutant has a spectral shift of only -0.12 eV ($\sim$ 40 %) experimentally. The calculations predict -0.19 eV ($\sim$ 60 %) for OM2/MRCI and -0.16 eV ($\sim$ 50 %) for SORCI. Thus, the calculations reproduce but overestimate the effect. The computed OM2/MRCI shift agrees nicely with the value obtained in the perturbation analysis (Table 6.8) when the side chains of all polar residues of the binding pocket are removed simultaneously (-0.20 eV).

Furthermore, the results are consistent with a study [324], in which the helices D, E, and G were identified as the determining factors for color regulation accounting for around 80 % of the spectral shift. Almost all residues with sizable impact on the excitation energies (Table 6.8) found in the present study are located on these helices.

6.5.6 Vibrational Analysis

Vibrational properties of rhodopsins have been widely used to elucidate various issues in bR and ppR, e.g., structures and structural changes, the mechanism of proton transport in bR and, in particular, the mechanism of the spectral tuning [402, 407, 426, 434]. In this work, two aspects of the vibrational characteristics of the chromophore backbone and of the Schiff base were studied for bR, ppR, and the bR/ppR mutant.

The NH stretching mode of the Schiff base is of particular interest because it reflects the strength of the hydrogen bond of the Schiff base. By performing QM/MM molecular dynamics simulations, the N-H stretching frequency was calculated in bR, ppR and the bR/ppR mutant via the velocity autocorrelation function.

In going from bR to ppR, this frequency is found shifted to lower wavenumbers by 25 cm^{-1} (Table 6.10), indicating a stronger hydrogen bond of the Schiff base to W402 in ppR. The calculated shift compares well to the experimentally observed one of 32-33 cm^{-1} [407] for the N-D stretching vibration of the deuterated Schiff base. For the bR/ppR mutant, a similar frequency ($\Delta\nu = 4$ cm^{-1}) is found as for ppR (Table 6.10), implying a similar hydrogen bond strength. This agrees well with experimental results [323], which demonstrate that the Schiff base vibrations remain ppR-like while the chromophore bands are altered from ppR-like to bR-like.

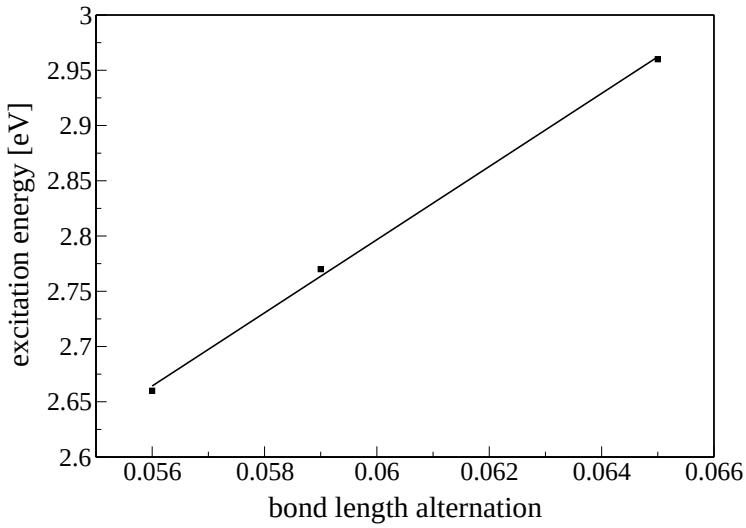


Figure 6.11: Correlation of bond length alternation and excitation energy for bR, ppR and the bR/ppR mutant.

The vibrations of the chromophore backbone, in particular the C=C stretching modes, are a sensitive probe for the electrostatic interaction of the retinal with the protein environment, especially for the potential gradient along the polyene chain [41, 42, 435, 436]. It is well-known that the C=C stretching frequency correlates with the bond length alternation (BLA) [41, 435, 437], which increases from bR (0.056 Å) via the bR/ppR mutant (0.059 Å) to ppR (0.065 Å). A similar increase is seen in the calculated excitation energies (Table 6.10). A linear relationship between the BLA and the excitation energies is found (Fig. 6.11), which is consistent with the experimentally characterized linear relationship between C=C stretching frequencies and the absorption maximum [41]. These results indicate that the electrostatic interaction with the protein environment is very similar in the bR/ppR mutant and in bR which share the same binding pocket.

6.5.7 Dynamics

To investigate the effects of structural fluctuations at room temperature on various properties of the systems, in particular on the stability of the HBN, QM/MM molecular dynamics simulations were performed for both proteins. As an approximation to the absorption spectra, histograms of the excitation energy (Fig. 6.12) along the trajectories [55] were calculated. A more profound line shape analysis would be preferable but is beyond the scope of the present study. Absorption maxima λ_{max} of 2.65 and 2.97 eV were found for bR and ppR, respectively, using Gaussian fits, thus reproducing the spectral shift between bR and ppR (0.32 eV; exp. 0.32 eV). The calculated line width of 2097 cm^{-1} (0.26 eV) for bR is comparable to the experimental one of 2860 cm^{-1} [233].

For the geometrical parameters of the extended HBN, fluctuations of about 0.2 Å (Table 6.11) around the equilibrium values are found thus indicating a fairly rigid

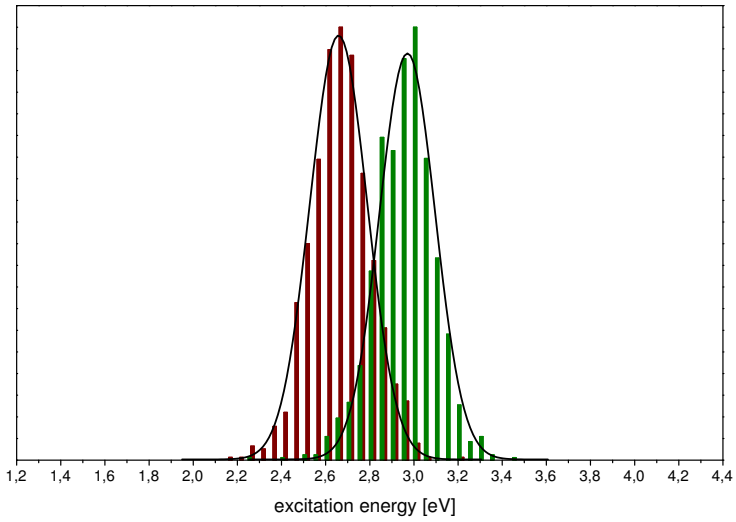


Figure 6.12: Calculated histograms for the excitation energy of the S_0-S_1 transition for bR (red) and ppR (green).

structure of the HBN. This stability of the HBN and of its interaction with the chromophore and counterion complex is a necessary requirement for applying a static approach that draws conclusions from the properties of optimized QM/MM structures.

The dynamics runs allow further to analyze the influence of structural fluctuations of the protein [438] on the excitation energy, by calculating the correlation coefficients between various geometrical parameters and the excitation energy along the MD trajectories. The results show a strong correlation (0.83) between the BLA and the excitation energy (Fig. 6.13), again consistent with the linear relationship between the excitation energy and the C=C stretching vibration of the chromophore [41] (Fig. 6.11) described in the preceding section.

The correlation of the excitation energy with individual distances between the Schiff base and the counterion residues (e.g., N_ζ-C_γ) is unexpectedly small (less than 0.1). The electrostatic potential generated by the partial charges on the counterion residues at the position of the Schiff-base nitrogen provides a collective coordinate that takes into account the correlated fluctuation of the counterion residues, which are not described using only single decoupled distance measures. A noticeable correlation (-0.3) is found between this potential and the excitation energy. Thus, the (collective) distance of the counterions from the chromophore influences the excitation energies, but this correlation is considerably weaker than that for the BLA.

6.6 Discussion

Several experimental [27, 324, 325] and theoretical investigations [50, 51] have addressed the spectral shift between bR and ppR, examining the different mechanisms of color tuning (i) - (iii) outlined in above. The calculations in this work allow for a distinc-

Table 6.11: Selected geometrical parameters and their fluctuations of the retinal binding site for QM/MM molecular dynamic simulations [a]. (Distances in Å; the two residue numbers specified refer to bR and ppR, respectively; fluctuations in parentheses)

		bR	ppR
RET_{N_ζ}	O(W402/402)	2.88 (0.12)	2.87 (0.13)
O(W402/402)	$O_{\delta 2}$(Asp85/75)	2.62 (0.10)	2.63 (0.10)
O(W402/402)	$O_{\delta 2}$(Asp212/201)	2.90 (0.23)	2.79 (0.16)
RET_{N_ζ}	$O_{\delta 1}$(Asp85/75)	3.81 (0.29)	3.73 (0.32)
RET_{N_ζ}	$O_{\delta 2}$(Asp85/75)	4.31 (0.30)	4.26 (0.35)
RET_{N_ζ}	C_γ(Asp85/75)	4.43 (0.24)	4.38 (0.29)
RET_{N_ζ}	$O_{\delta 1}$(Asp212/201)	4.07 (0.32)	4.43 (0.32)
RET_{N_ζ}	$O_{\delta 2}$(Asp212/201)	4.88 (0.26)	4.79 (0.26)
RET_{N_ζ}	C_γ(Asp212/201)	4.37 (0.23)	4.50 (0.25)
O(W402/402)	$O_{\delta 1}$(Asp85/75)	3.70 (0.27)	3.65 (0.26)
O(W402/402)	$O_{\delta 1}$(Asp212/201)	3.53 (0.25)	3.79 (0.26)
O(W401/401)	$O_{\delta 2}$(Asp85/75)	2.65 (0.11)	2.63 (0.10)
O(W401/401)	O(W406/400)	2.76 (0.13)	2.80 (0.15)
O(W406/400)	$O_{\delta 2}$(Asp212/212)	2.64 (0.10)	2.75 (0.25)
O(W406/400)	$N_{\omega/\delta}$(Arg82/72)	2.80 (0.11)	3.04 (0.25)
RET_{N_ζ}	C_ζ(Arg82/72)	10.42 (0.23)	10.33 (0.28)
RET_{C6}	C_ζ(Arg82/72)	11.72 (0.24)	10.95 (0.30)

[a] From QM calculations on the chromophore.

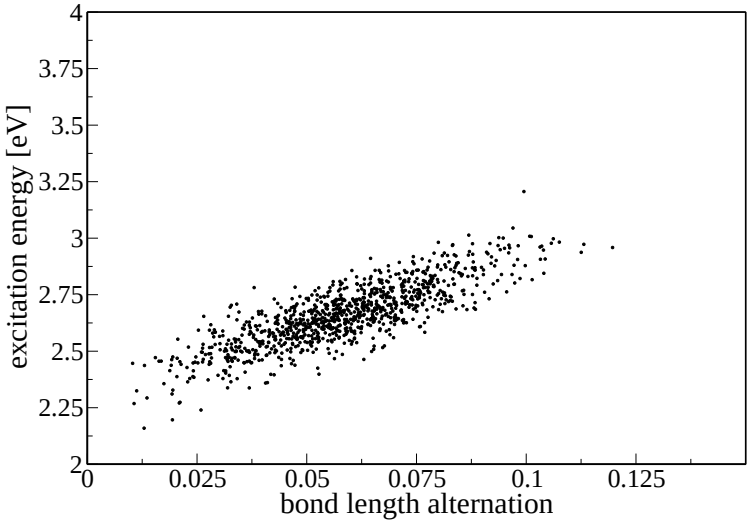


Figure 6.13: Correlation of bond length alternation and excitation energy for bR.

tion between these factors and a quantitative evaluation of their individual contribution.

(i) The differences between the chromophore structures of bR and ppR are already small in the crystal structures and nearly removed by the present QM/MM geometry optimizations, in agreement with previous theoretical studies [50, 51]. The remaining difference concerns essentially only the BLA and gives rise to a difference in excitation energies of 0.05 eV at the SORCI level of theory and 0.02 eV with OM2/MRCI, which is only a small contribution to the shift of the absorption maximum [††]. Even when using the nonoptimized X-ray structures, the difference in the OM2/MRCI excitation energies is not larger.

(ii) The polar and charged groups of the protein environment in bR and ppR interact differently with the charge density distribution on the chromophore in its ground and excited state, thereby stabilizing or destabilizing the respective state and so modulating the excitation energy. This polarization of the chromophore is reflected in the bond length alternation along the polyene chain [‡‡] as well as in the change of the dipole moment (with respect to vacuum). Both effects are seen in the present study. The larger bond length alternation in ppR as well as the larger change of the dipole moment due to the protein interaction ($|\mu_{S_0}^{\text{protein}} - \mu_{S_0}^{\text{vacuum}}|$) indicate a stronger electrostatic interaction with the protein in ppR compared to bR.

Furthermore, the vibrations of the chromophore backbone, in particular the C=C stretching modes, probe the electrostatic interaction of the polarized chromophore with the protein environment very sensitively [41, 42, 435, 436]. Using the bond length alternation as a measure for the C=C stretching frequencies [41], the calculations for bR, ppR, and the bR/ppR mutant (a multiple mutant in which the ten different amino acids of the binding pocket of ppR are mutated to their corresponding residues in bR) reproduce an experimentally observed linear correlation between the C=C stretching frequencies and the absorption maximum [39, 41, 439] very well. The comparison of bR, ppR, and the bR/ppR mutant demonstrates clearly that the chromophore vibrations in bR/ppR are bR-like, in agreement with experiment [323]. This points to a very similar electrostatic interaction of the chromophore with the protein in bR and this mutant, which means that this interaction is dominated by the binding pocket.

In addition to these global aspects, the results of the perturbation analysis and mutation experiments provide a detailed picture of the interaction of individual polar and charged groups of the environment with the chromophore. In agreement with experimental data, several sites (e.g., Thr204Ala, Gly130Ser, Val108Met, Ala131Thr) located in the binding pocket have a significant impact, however, none of them is solely responsible for the complete shift. Moreover, the perturbation analysis reveals a small but distinct influence of some sites (e.g., Ala111Thr, Ser44Pro), which do not belong to the binding pocket and have not been experimentally studied yet.

Nevertheless, the interactions with polar groups account only for a fraction, albeit

[††]Using the chromophore structures obtained by QM/MM optimization of the whole protein, a distinction between distortions induced by steric confinement and effects induced by electrostatic interactions with the protein is not possible.

[‡‡]It has been shown that the bond length alternation of the chromophore increases in the presence of a counterion close to the Schiff base [61, 410].

a major fraction, of the observed spectral shift. The analysis of multiple mutants emphasizes this fact. The bR/ppR mutant [27] shows experimentally only 44 % of the spectral shift; our calculations yield ∼50 %, in good agreement with experiment. These results confirm that none of the residues in the binding pocket is solely responsible for the spectral shift.

(iii) The interaction of the counterion residues with the chromophore strongly depends on their distance [25, 39] from the Schiff base but also on other charged and polar groups that are connected to the HBN.

Therefore, this often discussed mechanism of color tuning focuses on the structure of the counterion complex and the three water molecules connecting the Schiff base and the counterion residues (i.e., the extended HBN). The structural differences between bR and ppR mainly concern the different orientation of the guanidinium group of Arg82/72 and a distortion of the pentagonal cluster formed by the two aspartic acids of the counterion complex and the three water molecules [426] in ppR.

The contribution of the extended HBN to the spectral shift between bR and ppR have been estimated by including only the residues of the extended HBN as point charges in the excited state calculations, excluding the remainder of the protein. For this extended HBN model, a contribution of 23 % to the spectral shift is found. This value agrees well with the results of the perturbation analysis.

Including further residues from the extracellular site (Glu194 for bR, Asp193 for ppR) in the perturbation analysis yields a contribution of the complete hydrogen bonded network at the extracellular side of ∼30 % to the spectral shift.

This value seems to be a lower bound because the QM/MM minimized structures do not distort the pentagonal cluster in ppR to the extent that is found experimentally. When calculating the excitation energy with partially optimized structures (constrained to conserve the HBN positions from the crystal structure), a contribution of 40 % to the spectral shift is obtained for the counterion complex, instead of 23 % with the fully optimized structures. Therefore considering the distortion of the HBN and further residues from the extracellular site, the contribution of the counterion complex to the spectral shift may be as high as 40 %.

The interaction of the counterion complex with the chromophore also affects the vibrations of the Schiff base. For instance, the NH stretching mode of the Schiff base reflects the strength of the hydrogen bond of the Schiff base to the counterion complex. The calculations in this study yield a clearly increased frequency for bR compared to ppR whereas it is essentially unchanged (even lower) in the bR/ppR mutant. This suggests, in agreement with the experiment, a stronger hydrogen bond of the Schiff base with the counterion complex in ppR which is hardly influenced by the mutations in the binding pocket that are present in the bR/ppR mutant.

The stronger hydrogen bond in ppR was supposed to originate from small changes in the HBN and related to the spectral blue shift in ppR [323, 402]. Results from X-ray crystallography [332, 378] and the analysis of the O-D stretching vibrations of the water molecules in the HBN [407, 426] support a distorted HBN structure in ppR.

Despite the rather similar structure of the HBN in our QM/MM optimized structures (sec. 6.5.2), a stronger hydrogen bond between the Schiff base and water W402 is observed. Different polarizations of the chromophore due to the protein environ-

ments also reflected in the dipole moments (Table 6.2) could be a factor explaining the result. A distortion of the HBN structure might be an additional factor, explaining why the contribution of the counterion complex is underestimated in this work.

In summary, the results indicate that the interaction with the extended HBN and with the binding pocket equally contribute to the spectral shift. Taking into account the effect of the distorted HBN, the influence of the extended counterion region (including Glu204 for bR and Asp193 for ppR) is estimated to cause about ∼40 % of the spectral shift. The best estimate for the contribution of the binding pocket amounts to about 40-50 %.

These results explain why the experimental absorption maximum of the bR/ppR mutant is less shifted toward the bR value than expected [27]. Due to the mutations, the binding pocket of this mutant is bR-like, whereas the counterion region remains ppR-like. Thus the replacement of the amino acids within 5 Å of the chromophore is not sufficient to transform ppR into bR with regard to electronic excitation.

Chapter 7

Summary

The simulation of bioenergetical processes such as proton transfer reactions or photoexcitations is a challenging task even for modern simulation techniques. The goal of this work was to investigate two interesting problems of molecular bioenergetics – the water/proton selectivity in aquaporins and the color tuning in rhodopsins – using multiple-length-scale approaches.

Development of a new Reaction Coordinate for PT and a Model Channel

A new reaction coordinate was introduced for characterizing the progress of long-range PT. The new reaction coordinate eliminates the problems encountered by earlier suggestions which were also analyzed here, works without assuming a priori the pathway, and a generalization to three-dimensional transfer pathways is, in contrast to earlier suggestions, easily possible as demonstrated in a joint publication [130].

The new coordinate can be used as the reaction coordinate for computing meaningful potential of mean force for long-range PT, as shown for the PT in a model channel and for other systems, such as carbonic anhydrase [130].

Further, the influence of electrostatics in long-range PT was also briefly investigated with PT in the model channel embedded in different dielectric continuum environments. The results indicate that PT through a purely hydrophobic channel embedded in a membrane encounters a significant barrier, unless polar interactions are present in the channel to stabilize the hydronium ion.

Proton Blockage in GlpF

Employing a QM/MM/continuum electrostatics simulation protocol and the new reaction coordinate, the proton blockage in the aquaporin GlpF was investigated.

The calculation of the free energy barrier impeding the PT through GlpF yields a barrier of $\sim$25 kcal/mol which is sufficiently high to prevent PT through the pore. A perturbation analysis determined the contribution of specific element to the free energy barrier. The results indicate that the main contribution to the free energy barrier is the desolvation penalty for transferring an excess proton from bulk solvent to the single water file through the pore. Structural motifs such as the NPA motifs are of minor importance but of course influence the barrier. The central topic is therefore

the question about the importance of structural elements versus the reference state. This topic is also reflected in the controversial effect of the Arg of the SF. Whereas the energetical analysis in this work indicates a minor influence on the overall barrier, another study [112] suggest a significant influence. However, in this study the energetical influence is admixed with a change of the reference point caused by changed desolvation effects.

Furthermore, the influence of polar groups in the pore on the stabilization of an excess proton clarified that the real challenge is to design a channel with correct polarity to allow proton or more generally ion conduction. As the results for GlpF and the model channel demonstrated, the reverse, the construction of a channel that prevents proton or ion conduction, is easy to achieve. Due to the desolvation penalty of the excess proton or ion, a relatively hydrophobic channel with a diameter not to large is sufficient.

Color Tuning in Archaeal Rhodopsins

As second application, the mechanism of color tuning in archaeal rhodopsins was studied by comparing the optical properties of the light-driven proton pump bacteriorhodopsin (bR) and the light detector phoborhodopsin (ppR).

The calculated shift of the absorption maximum λ_{max} of ppR relative to bR as well as the magnitude of the bandwidth agree well with experimental results. Mutation studies and the analysis of vibrational properties indicate that numerous sources contribute to the spectral shift between bR and ppR. The two main and equally important factors, which are responsible for about 90 % of the total shift, are the different neutral amino acids in the binding pocket and the difference in the extended hydrogen bonded network at the extracellular side of the proteins. The rest may be accounted for by geometrical differences between the chromophores and the conserved, polarizable residues, i.e., by slight differences in their electrostatic interaction, including protein polarization and differential dispersive interactions.

The Investigation of Bioenergetical Processes with Multiple-Length-Scale Approaches

Both applications and the simulation of the model channel demonstrated the ability of multiple-length-scale approaches to simulate various bioenergetical processes. Their application is often indispensable to reduce the prohibitive computational cost connected with their simulation.

For the model channel, the effect of environments with variable electrostatic properties on the PT energetics was successfully demonstrated. The application of the employed QM/MM/continuum electrostatics simulation protocol is therefore justified for incorporating this aspect of heterogeneous environments found in biological systems; and the basis for the following simulation of the proton blockage in GlpF was established. The comparison of the QM/MM with full QM results further showed that the usage of a QM/MM approach to simulate PT trough channels in a complex dielectric environment is well-founded.

The simulation of the proton blockage in GlpF required the use of dynamical methods in order to properly describe the solvent reorganization during the PT. The QM/MM/continuum electrostatics simulation protocol thereby turned out to be an

attractive alternative to reduce the prohibitive computational cost connected with the use of dynamical methods. More generally, it was shown that slab models in conjunction with GSBP can be used to describe fundamental processes in (membrane) proteins, saving computer time else needed for explicit representation and nonetheless reproduce the crucial aspects of huge simulation setups.

The study of the color tuning in rhodopsins underlined that combining very efficient (SCC-DFTB, OM2/MRCI) and very accurate (SORCI) methods in multiple-length-scale approaches allows a comprehensive, qualitative investigation of the spectral shift including the analysis of dynamical effects. The employed methodology even correctly describes the subtle effects of point and multiple mutations. So the properties of bR, ppR, and the bR/ppR mutant in the Schiff base region, illustrated by the N-H stretching vibration, as well as in the binding pocket, exemplified, e.g., by the experimentally observed linear correlation between the C=C stretching frequencies of the chromophore and the absorption maximum [39, 41, 42], were correctly described.

The good agreement between theoretical and experimental results for the color tuning in archaeal rhodopsins and the water structure and proton blockage in GlpF shows that modern quantum mechanical methods and multiple-length-scale approaches can not only reproduce but also interpret bioenergetical phenomena.

Bibliography

[1] Lohmann, K. *Naturwissenschaften* **1929**, *17*, 624.

[2] Fiske, C. H.; SubbaRow, Y. *Science* **1929**, *70*, 381.

[3] Lipmann, F. *Advances in Enzymology* **1941**, *1*, 99.

[4] Mitchell, P. *Nature (London)* **1961**, *191*, 144.

[5] Stoeckenius, W.; Lozier, R. H.; Bogomolni, R. A. *Biochim. Biophys. Acta* **1979**, *505*, 215.

[6] Oren, A. *FEMS Microbiol. Rev.* **1994**, *13*, 415.

[7] Mathies, R. A.; Lin, S. W.; Ames, J. B.; Pollard, W. T. *Annu. Rev. Biophys. Biophys. Chem.* **1991**, *20*, 491.

[8] Borgnia, M.; Nielsen, S.; Engel, A.; Agre, P. *Annu. Rev. Biochem.* **1999**, *68*, 425.

[9] van der Horst, M. A.; Hellingwerf, K. J. *Acc. Chem. Res.* **2004**, *37*, 13.

[10] Spudich, J. L.; Yang, C.; Jung, K.; Spudich, E. N. *Annu. Rev. Cell Dev. Biol.* **2000**, *16*, 365.

[11] Schwemer, J.; Langer, H. *Methods Enzymol* **1982**, *81*, 182.

[12] Yokoyama, S. *Prog. Retinal Eye Res.* **2000**, *19*, 419.

[13] Lythgoe, J. N. *Vision Res.* **1968**, *8*, 997.

[14] Yokoyama, S.; Yokoyama, R. *Annu. Rev. Ecol. Syst.* **1996**, *27*, 543.

[15] Nathans, J. *Annu. Rev. Neurosci.* **1987**, *10*, 163.

[16] Oprian, D. D.; Asenjo, A. B.; Lee, N.; Pelletier, S. L. *Biochemistry* **1991**, *30*, 11367.

[17] Wald, G.; Brown, P. K. *J. Gen. Physiol.* **1953**, *37*, 189.

[18] Ollivier, B.; Caumette, P.; Garcia, J. L.; Mah, R. A. *Microbiol Mol Biol Rev.* **1994**, *58*, 27.

[19] Schaefer, G.; Engelhardt, M.; Mueller, V. *Microbiol. Mol. Biol. Rev.* **1999**, *63*, 570.

[20] Chizhov, I.; Schmies, G.; Seidel, R.; Sydor, J. R.; Lüttenberg, B.; Engelhard, M. *Biophys. J.* **1998**, *75*, 999.

[21] Nakanishi, K.; Balogh-Nair, Y.; Arnaboldi, M.; Tsujimoto, K.; Honig, B. *J. Am. Chem. Soc.* **1980**, *102*, 7945.

[22] Andersen, L. H.; Nielsen, I. B.; Kristensen, M. B.; El Ghazaly, M. O. A.; Haacke, S.; Nielsen, M. B.; Petersen, M. A. *J. Am. Chem. Soc.* **2005**, *127*, 12347.

[23] Spudich, J. L.; McCain, D. A.; Nakanishi, K.; Okabe, M.; Shimizu, N.; Rodman, H.; Honig, B.; Bogomolni, R. A. *Biophys. J.* **1986**, *49*, 479.

[24] Takahashi, T.; Yan, B.; Mazur, P.; Derguini, F.; Nakanishi, K.; Spudich, J. L. *Biochemistry* **1990**, *29*, 8467.

[25] Hu, J.; Griffin, R. G.; Herzfeld, J. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 8880–8884.

[26] Lin, S. W.; Kochendoerfer, G. G.; Carroll, K. S.; Wang, D.; Mathies, R. A.; Sakmar, T. P. *J. Biol. Chem.* **1998**, *273*, 24583.

[27] Shimono, K.; Iwamoto, M.; Sumi, M.; Kamo, N. *Biochim. Biophys. Acta* **2001**, *1515*, 92.

[28] Parry, J. W. L.; Poopalasundaram, S.; Bowmaker, J. K.; Hunt, D. M. *Biochemistry* **2004**, *43*, 8014.

[29] Kochendoerfer, G. G.; Lin, S. W.; Sakmar, T. T.; Mathies, R. A. *TIBS* **1999**, *24*, 300–305.

[30] Yan, B.; Spudich, J. L.; Mazur, P.; Vunnam, S.; Derguini, F.; Nakanishi, K. *J. Biol. Chem.* **1995**, *270*, 29668.

[31] Chan, T.; Lee, M.; Sakmar, T. P. *J. Biol. Chem.* **1992**, *267*, 9478.

[32] Okada, T.; Sugihara, M.; Bondar, A.; Elstner, M.; Entel, P.; Buss, V. *J. Mol. Biol.* **2004**, *342*, 571.

[33] Kosower, E. M. *Proc. Natl. Acad. Sci., USA* **1988**, *85*, 1076.

[34] Williams, A. J.; Hunt, D. M.; Bowmaker, J. K.; Mollon, J. D. *EMBO J.* **1992**, *11*, 2039.

[35] Yokoyama, R.; Yokoyama, R. *Proc. Natl. Acad. Sci., USA* **1990**, *87*, 9315.

[36] Neitz, M.; Neitz, J.; Jacobs, G. H. *Science* **1991**, *252*, 971.

[37] Baasov, T.; Sheves, M. *J. Am. Chem. Soc.* **1985**, *107*, 7524.

[38] Blatz, P. E.; Mohler, J. H.; Navangul, H. V. *Biochemistry* **1972**, *11*, 848.

[39] Baasov, T.; Friedman, N.; Sheves, M. *Biochemistry* **1987**, *26*, 3210–3217.

[40] Fodor, S. P. A.; Gebhard, R.; Lugtenburg, J.; Bogomolni, R. A.; Mathies, R. A. *J. Biol. Chem.* **1989**, *264*, 18280.

[41] Kochendoerfer, G.; Wang, Z.; Oprian, D. D.; Mathies, R. A. *Biochemistry* **1997**, *36*, 6577–6587.

[42] Lin, S.; Imamoto, Y.; Fukada, Y.; Shichida, Y.; Yoshizawa, T.; Mathies, R. *Biochemistry* **1994**, *33*, 2151.

[43] Blatz, P. E.; Mohler, J. H. *Biochemistry* **1975**, *14*, 2304.

[44] Honig, B.; Hudson, B.; Sykes, B. D.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **1971**, *68*, 1289.

[45] Harbison, G. S.; Mulder, P. P. J.; Pardoen, J. A. *Biochemistry* **1985**, *24*, 6955.

[46] De Groot, H. J. M.; Smith, S. O.; Courtin, J.; Van den Berg, E.; Winkel, C.; Lugtenburg, J.; Griffin, R. G.; Herzfeld, J. *Biochemistry* **1990**, *29*, 6873.

[47] Albeck, A.; Livnah, N.; Gottlieb, H.; Sheves, M. *J. Am. Chem. Soc.* **1992**, *114*, 2400.

[48] Asenjo, A. B.; Rim, J.; Oprian, D. D. *Neuron* **1994**, *12*, 1131.

[49] Honig, B.; Greenberg, A. D.; Dinur, U.; Ebrey, T. G. *Biochemistry* **1976**, *15*, 4593.

[50] Ren, L.; Martin, C. H.; Wise, K. J.; Gillespie, N. B.; Luecke, H.; Lanyi, J. K.; Spudich, J. L.; Birge, R. R. *Biochemistry* **2001**, *40*, 13906–13914.

[51] Hayashi, S.; Tajkhorshid, E.; Pebay-Peyroula, E.; Royant, A.; Landau, E. M.; J., N.; Schulten, K. *J. Phys. Chem. B* **2001**, *105*, 10124–10131.

[52] Houjou, H.; Inoue, Y.; Sakurai, M. *J. Am. Chem. Soc.* **1998**, *120*, 4459–4470.

[53] Houjou, H.; Inoue, Y.; Sakurai, M. *J. Phys. Chem. B* **2001**, *105*, 867–879.

[54] Houjou, H.; Koyama, K.; Wada, M.; Sameshima, K.; Inoue, Y.; Sakurai, M. *Chem. Phys. Lett.* **1998**, *294*, 162–166.

[55] Warshel, A.; Chu, Z. T. *J. Phys. Chem. B* **2001**, *105*, 9857–9871.

[56] Rajamani, R.; Gao, J. *J. Comput. Chem.* **2002**, *23*, 96.

[57] Ferre, N.; Olivucci, M. *J. Am. Chem. Soc.* **2003**, *125*, 6868.

[58] Kloppman, E.; Becker, T.; Ullmann, G. M. *Proteins* **2005**, *61*, 953.

[59] Kakitani, T.; Beppu, Y.; Yamada, A. *Photochem. Photobiol.* **1999**, *70*, 686.

[60] Beppu, Y. *J. Phys. Soc. Jpn* **1997**, *66*, 3303.

[61] Schreiber, M.; Buss, V.; Sugihara, M. *J. Chem. Phys.* **2003**, *119*(23), 12045–12048.

[62] Schreiber, M.; Buss, V. *Int. J. Quantum Chem.* **2003**, *95*, 882.

[63] Finkelstein, A. *Water Movement Through Lipid Bilayers, Pores, and Plasma Memebranes. Theory and Reality.;* Wiley: New York, 1987.

[64] Macey, R. I. *Am. J. Physiol. (Cell Physiol.)* **1984**, *246*, 195.

[65] Solomon, A. K. *J. Gen. Physiol.* **1968**, *51*, 335.

[66] Denker, B. M.; Smith, B. L.; Kuhajda, F. P.; Agre, P. *J. Biol. Chem.* **1988**, *263*, 15634.

[67] Smith, B. L.; Agre, P. *J. Biol. Chem.* **1991**, *266*, 6407.

[68] Preston, G. M.; Agre, P. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 11110.

[69] Preston, G. M.; Carroll, T. P.; Guggino, W. B.; Agre, P. *Science* **1992**, *256*, 385.

[70] Agre, P.; Bonhivers, M.; Borgnia, M. J. *J. Biol. Chem.* **1998**, *1998*, 14659.

[71] Agre, P.; King, L. S.; Yasui, M.; Guggino, W. B.; Ottersen, O. P.; Fujiyoshi, Y.; Engel, A.; Nielsen, S. *J. Physiol.* **2002**, *542*, 3.

[72] Nielsen, S.; Smith, B. L.; Christensen, E. I.; Knepper, M. A.; Agre, P. *J. Cell Biol.* **1993**, *120*, 371.

[73] Fushimi, K.; Uchida, S. *Nature (London)* **1993**, *361*, 549.

[74] Nielsen, S.; DiGiovanni, S. R.; Christensen, E. I.; Knepper, M. A.; Harris, H. W. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 11663.

[75] Schrier, R. W.; Cadnapaphornchai, M. A. *Prog. Biophys. Mol. Biol.* **2003**, *81*, 117.

[76] Deen, P. M. T.; Verdijk, M. A. J.; Knoers, N. V. A. M.; Wieringa, B.; Monnens, L. A. H.; van Os, C. H.; van Oost, B. A. *Science* **1994**, *264*, 92.

[77] Gorin, M. B.; Yancey, S. B.; Cline, J.; Revel, J.; Horwitz, J. *Cell* **1984**, *39*, 49.

[78] Berry, V.; Francis, P.; Kaushal, S.; Moore, A.; Bhattacharya, S. *Nat. Genet.* **2000**, *25*, 15.

[79] Francis, P.; Chung, J.; Yasui, M.; Berry, V.; Moore, A.; Wyatt, M. K.; Wistow, G.; Bhattacharya, S. S.; Agre, P. *Hum. Mol. Genet.* **2000**, *9*, 2329.

[80] Heller, K. B.; Lin, E. C. C.; Wilson, T. H. *J. Bactriol.* **1980**, *144*, 274.

[81] Borgnia, M.; Agre, P. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 2888.

[82] Fu, D.; Libson, A.; Miercke, J. W.; Weitzman, C.; Nollert, P.; Krucinski, J.; Stroud, R. M. *Sience* **2000**, *290*, 481.

[83] Grayson, P.; Tajkhorshid, E.; Schulten, K. *Biophys. J.* **2003**, *85*, 36.

[84] Marx, D.; Tuckerman, M. E.; Hutter, J.; Parrinello, M. *Nature (London)* **1999**, *397*, 601.

[85] Schmitt, U. W.; Voth, G. A. *J. Chem. Phys.* **1999**, *111*, 9361.

[86] Grotthus, C. *Annales de Chimie* **1806**, *LVIII*, 54.

[87] Agmon, N. *Chem. Phys. Lett.* **1995**, *244*, 456.

[88] Tajkhorshid, E.; Nollert, P.; Morten, J.; Miercke, L. J. W.; O'Connel, J.; Stroud, R. M.; Schulten, K. *Science* **2002**, *296*, 525.

[89] de Groot, B. L.; H., G. *Science* **2001**, *294*, 2353.

[90] Kong, Y.; Ma, J. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 14345.

[91] Pomès, R.; Roux, B. *Biophys. J.* **1996**, *71*, 19.

[92] Pomès, R.; Roux, B. *Biophys. J.* **2002**, *82*, 2304.

[93] Schumaker, M. F.; Pomès, R.; Roux, B. *Biophys. J.* **2000**, *79*, 2840.

[94] Stroud, R. M.; Miercke, L. J.; O'Connell, J. annd Khademi, S.; Lee, J. K.; Remis, J.; Harries, Y. R.; Akhavan, D. *Curr. Opin. Struct. Biol.* **2003**, *13*, 424.

[95] Sansom, M. S. P.; Law, R. J. *Curr. Biol.* **2001**, *11*, R71.

[96] Savage, D. F.; Egea, P. F.; Robles-Colmenares, Y.; O'Connell, J. D.; Stroud, R. M. *PLoS Biol.* **2003**, *1*, 334.

[97] Sui, H.; Han, B.; Lee, J. K.; Walian, P.; Jap, B. K. *Science* **2001**, *414*, 872.

[98] Harries, W. E. C.; Akhavan, D.; Miercke, L. J. W.; Khademi, S.; Stroud, R. M. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 14045.

[99] Gonen, T.; Sliz, P.; Kistler, J.; Cheng, Y.; Walz, T. *Nature (London)* **2004**, *429*, 193.

[100] Ren, G.; Reddy, V. S.; Cheng, A.; Melnyk, P.; Mitra, A. K. *Proc. Nat. Acad. Sci.* **2001**, *98*, 1398.

[101] Murata, K.; Mitsuoka, K.; Hirai, T.; Walz, T.; Agre, P.; Heymann, J. B.; Engel, A.; Fujiyoshi, Y. *Nature (London)* **2000**, *407*, 599.

[102] Beitz, E.; Wu, B.; Hom, L. M.; Schultz, J. E.; Zeuthen, T. *Proc. Nat. Acad. Sci.* **2006**, *103*, 269.

[103] Lui, K.; Kozono, D.; Kato, Y.; Agre, P.; Hazama, A.; Yasui, M. *Proc. Nat. Acad. Sci.* **2005**, *102*, 2193.

[104] Miloshevsky, G. V.; Jordan, P. C. *Biophys. J.* **2004**, *87*, 3690.

[105] Morten, J.; Tajkhorshid, E.; Schulten, K. *Biophys. J.* **2003**, *85*, 2884.

[106] Chakrabarti, N.; Tajkhorshid, E.; Roux, B.; Pommes, R. *Structure* **2004**, *12*, 65.

[107] Chakrabarti, N.; Roux, B.; Pommes, R. *J. Mol. Biol.* **2004**, *343*, 493.

[108] de Groot, B. L.; Frigato, T.; Helms, V.; Grubmüller, H. *J. Mol. Biol.* **2003**, *333*, 279.

[109] Burykin, A.; Warshel, A. *Biophys. J.* **2003**, *85*, 3696.

[110] Ilan, B.; Tajkhorshid, E.; Schulten, K.; Voth, G. A. *Protein* **2004**, *55*, 223.

[111] Burykin, A.; Warshel, A. *FEBS Lett.* **2004**, *570*, 41.

[112] Chen, H.; Wu, Y.; Voth, G. A. *Biophys. J.* **2006**, *90*, L73.

[113] Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem.* **1990**, *11*, 700.

[114] Aaqvist, J.; Warshel, A. *Chem. Rev.* **1993**, *93*, 2523.

[115] Gao, J. *Acc. Chem. Res.* **1996**, *29*, 298.

[116] Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227.

[117] Gao, J.; Truhlar, D. G. *Annu. Rev. Phys. Chem.* **2002**, *53*, 467.

[118] Warshel, A. *Annu. Rev. Biophys. Biomol. Struct.* **2002**, *32*, 425.

[119] Cui, Q.; Karplus, M. *Adv. Prot. Chem.* **2003**, *66*, 315.

[120] Lennartz, C.; Schaefer, A.; Terstegen, F.; Thiel, W. *J. Phys. Chem. B* **2002**, *106*, 1758.

[121] Guallar, V.; Friesner, R. A. *J. Am. Chem. Soc.* **2004**, *126*, 8501–8508.

[122] Cisneros, G. A.; Liu, H. Y.; Zhang, Y. K.; Yang, W. T. *J. Am. Chem. Soc.* **2003**, *125*, 10384.

[123] Fersht, A. *Enzyme structure, mechanism and protein foldings.;* W. H. Freeman: New York, 1998.

[124] Honig, B.; Nicholls, A. *Science* **1995**, *268*, 1144.

[125] Warshel, A.; Papazyan, A. *Curr. Opin. Struct. Biol.* **1998**, *8*, 211.

[126] Warshel, A. *Annu. Rev. Biophys. Biomol. Struct.* **2003**, *32*, 425.

[127] Li, G.; Zhang, X.; Cui, Q. *J. Phys. Chem. B* **2003**, *107*, 8643.

[128] Schaefer, P.; Riccardi, D.; Cui, Q. *J. Chem. Phys.* **2005**, *123*, 014905.

[129] D. Riccardi, D.; Schaefer, P.; Cui, Q. *J. Phys. Chem. B* **2005**, *109*, 17715.

[130] König, P. H.; Ghosh, N.; Hoffmann, M.; Elstner, M.; Tajkhorshid, E.; Frauenheim, T.; Cui, Q. *J. Phys. Chem. A* **2006**, *110*, 548.

[131] Garcia-Viloca, M.; Truhlar, D. G.; Gao, J. *J. Mol. Biol.* **2003**, *327*, 549.

[132] Roux, B. In *Computational Biochemistry and Biophysics.*; Becker, O. M., MacKerell, A. D., Roux, B., Watanabe, M., Eds.; Marcel Dekker, Inc., 2001.

[133] Cramer, C. J.; Truhlar, D. G. *Chem. Rev.* **1999**, *99*, 2161.

[134] Beglov, D.; Roux, B. *J. Chem. Phys.* **1994**, *100*, 9050.

[135] Im, W.; Berneche, S.; Roux, B. *J. Chem. Phys.* **2001**, *114*, 2924.

[136] Cui, Q.; Elstner, M.; Kaxiras, E.; Frauenheim, T.; Karplus, M. *J. Phys. Chem. B* **2001**, *105*, 569.

[137] MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher III, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586.

[138] Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. *Phys. Rev. B* **1998**, *58*, 7260–7268.

[139] Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. E.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.

[140] Elstner, M.; Frauenheim, T.; Suhai, S. *J. Mol. Struct.* **2003**, *632*, 29–41.

[141] Elstner, M.; Cui, Q.; Munih, P.; Kaxiras, E.; Frauenheim, T.; Karplus, M. *J. Comput. Chem.* **2003**, *24*, 565.

[142] Zhang, X.; Harrison, D. H. T.; Cui, Q. *J. Am. Chem. Soc.* **2002**, *124*, 14871.

[143] Formaneck, M. S.; Li, G.; Zhang, X.; Cui, Q. *J. Theor. Comput. Chem.* **2002**, *1*, 53.

[144] Cui, Q.; Elstner, M.; Karplus, M. *J. Phys. Chem. B* **2002**, *106*, 2721.

[145] Li, G.; Cui, Q. *J. Am. Chem. Soc.* **2003**, *125*, 15028.

[146] Li, G.; Cui, Q. *J. Phys. Chem. B* **2003**, *107*, 14521.

[147] Weber, W.; Thiel, W. *Theor. Chem. Acc.* **2000**, *103*, 495–506.

[148] Strodel, P. *Beschreibung angeregter Molekülzustände in komplex strukturierten Umgebungen durch einen effizienten, individuell selektierenden MRCI-Algorithmus gekoppelt an ein molekularmechanisches Kraftfeld.* PhD thesis, Ludwig-Maximilians-Universität München, **2003**.

[149] Jensen, F. *Introduction to computational chemistry.;* John Wiley & Sons, 1999.

[150] Knowles, P.; Schütz, M.; Werner, H.-J. In *Modern Methods and Algorithms of Quantum Chemistry*, Vol. 3; John von Neumann Institute for Computing: Jülich, 2000; pages 97–179.

[151] Leach, A. R. *Molecular Modelling. Principles and Application;* Addison Wesley Longman Limited: Harlow, 1996.

[152] Szabo, A.; Ostlund, N. S. *Modern Quantum Chemistry;* Dover Publications, Inc.: Mineola, New York, 1996.

[153] Helgaker, T.; Jørgensen, P.; Olsen, J. *Molecular Electronic-Structure Theory;* John Wiley & Sons, 2000.

[154] Parr, R. G.; Yang, W. *Density-Functional Theory of Atoms and Molecules.;* Oxford University Press: New York, 1994.

[155] Hohenberg, P.; Kohn, W. *Phys Rev. B* **1964**, *136*, 864–871.

[156] Jones, R. O.; Gunnarsson, O. *Rev. Mod. Phys.* **1989**, *61*, 689.

[157] Kohn, W.; J., S. L. *Phys. Rev.* **1965**, *140*, 1133.

[158] Frauenheim, T.; Weich, F.; Köhler, T.; Uhlmann, S.; Porezag, D.; Seifert, G. *Phys. Rev. B* **1995**, *52*, 11492.

[159] Harrison, W. *Electronic Structure and the Properties of Solids.;* W.H. Freeman and Company: San Francisco, 1980.

[160] Frauenheim, T.; Seifert, G.; Elstner, M.; Niehaus, T.; Köhler, C. *J. Phys. Cond. Matt.* **2002**, *14*, 3015.

[161] Foulkes, W.; Haydock, R. *Phys. Rev. B* **1989**, *39*, 12520.

[162] Frauenheim, T.; Seifert, G.; Elstner, M.; Hajnal, Z.; Jungnickel, G.; Porezag, D.; Suhai, S.; Scholz, R. *physica status solidi (b)* **2000**, *217*, 41.

[163] Slater, J. C. *Phys. Rev.* **1930**, *35*, 210.

[164] Roothaan, C. C. J. *Rev. Mod. Phys.* **1951**, *23*, 69.

[165] Dewar, M. S.; Thiel, W. *J. Am. Chem. Soc.* **1977**, *99*, 4899.

[166] Dewar, M. S.; Zoebisch, E. G.; Healy, E. F.; Steward, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902.

[167] Steward, J. J. P. *J. Comput. Chem.* **1989**, *10*, 209.

[168] Pople, J. A.; Santry, D. P.; Segal, G. A. *J. Chem. Phys.* **1965**, *43*, S129.

[169] Pople, J. A.; Beveridge, D. L. *Approximate molecular orbital theory.;* McGraw-Hill: New York, 1970.

[170] Segal, G. A. *Semiempirical Methods of Electronic Structure Calculations.;* Plenum Press: New York, 1977.

[171] Parr, R. G. *J. Chem. Phys.* **1952**, *20*, 1499.

[172] Weinhold, F.; Carpenter, J. E. *J. Mol. Struct.* **1988**, *165*, 189.

[173] Cook, D. B.; Hollis, P. C.; McWeeny, R. *Mol. Phys.* **1967**, *13*, 553.

[174] Koch. *Z. Naturforsch. A* **1993**, *48*, 819.

[175] Weber, W. *Ein neues semiempirisches NDDO-Verfahren mit Orthogonalisierungskorrekturen.* PhD thesis, Universität Zürich, **1996**.

[176] Kolb, M.; Thiel, W. *J. Comput. Chem.* **1993**, *14*, 775.

[177] Hase, H.; Lauer, G.; Schulte, K.; Schweig, A. *Theoret. Chim. Acta* **1978**, *48*, 47.

[178] Borden, W. T.; Davidson, E. R. *Acc. Chem. Res.* **1996**, *29*, 67.

[179] Siegbahn, P. E. M. *Lecture notes in quantum chemistry.;* Springer: New York, 1992.

[180] Löwdin, P. O. *Int. J. Quantum Chem.* **1992**, *55*, 77.

[181] Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M. *J. Am. Chem. Soc.* **1992**, *114*, 10024.

[182] Allinger, N. L. *J. Am. Chem. Soc.* **1977**, *99*, 8127.

[183] Allinger, N. L.; Yuh, Y. H.; Lii, J. H. *J. Am. Chem. Soc.* **1989**, *111*, 8551.

[184] MacKerell, A. D. In *43th Sanibel Symposium*, 2003.

[185] Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E. I.; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. *Comput. Phys. Commun.* **1995**, *91*, 1.

[186] Ponder, J. W.; Case, D. A. *Adv. Protein Chem.* **2003**, *66*, 27.

[187] MacKerell, A. D.. In *Computational Biochemistry and Biophysics.*; Becker, O. M., MacKerell, A. D., Roux, B., Watanabe, M., Eds.; Marcel Dekker, Inc., 2001.

[188] van Gunsteren, W. F.; Billeter, S. R.; Eising, A. A.; Hünenberger, P. H.; Krüger, P.; Mark, A. E.; Scott, W. R. P.; Tironi, I. G. *Biomolecular Simulation: The GROMOS96 Manual and User Guide.;* BIOMOS: Zürich, 1996.

[189] Morse, P. M. *Phys. Rev.* **1929**, *34*, 57.

[190] Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.

[191] Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269.

[192] Chirlian, L. E.; Francl, M. M. *J. Comput. Chem.* **1987**, *8*, 894.

[193] Breneman, C. M.; Wiberg, K. B. *J. Comput. Chem.* **1990**, *11*, 361.

[194] Grubmüller, H. *Phys. Rev. E* **1995**, *52*, 2893.

[195] Voter, A. F. *Phys. Rev. Lett.* **1997**, *78*, 3908.

[196] Brooks, C. L.; Karplus, M.; Pettitt, B. M. *Proteins: A Theoretical Perspective of Dynamics, Structure, and Thermodynamics;* Wiley, 1990.

[197] Chandler, D. *Introduction to Modern Statistical Mechanics;* Oxford University Press, 1987.

[198] McQuarrie, D. A. *Statistical Mechanics;* HarperCollinsPublishers, 1976.

[199] Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. *J. Comput. Chem.* **1992**, *13*, 1011.

[200] Bartels, C.; Karplus, M. *J. Comput. Chem.* **1997**, *18*, 1450.

[201] Stewart, J. J. P. *J. Comput. Chem.* **1989**, *10*, 209.

[202] Weber, T. A.; Stillinger, F. H. *J. Chem. Phys.* **1982**, *77*, 4150.

[203] Riccardi, D.; Li, G.; Cui, Q. *J. Phys. Chem. B* **2004**, *108*, 6467.

[204] Svensson, M.; Humbel, S.; Froese, R. D. J.; Matsubara, T.; Sieber, S.; Morokuma, K. *J. Phys. Chem.* **1996**, *100*, 19357.

[205] Bakowies, D.; Thiel, W. *J. Phys. Chem.* **1996**, *100*, 10580.

[206] Sauer, J.; Sierka, M. *J. Comput. Chem.* **2000**, *21*, 1470–1493.

[207] Singh, U. C.; Kollmann, P. A. *J. Comput. Chem.* **1986**, *7*, 718.

[208] Sherwood, P.; de Vries, A. H.; Collins, S. J.; Greatbanks, S. P.; Burton, N. A.; A., V. M.; Hillier, I. H. *Faraday Discuss.* **1997**, *106*, 79.

[209] Lamoureux, G.; Roux, B. *J. Chem. Phys.* **2003**, *119*, 3025–3039.

[210] Thole, B. T. *Chem. Phys.* **1981**, *59*, 341.

[211] Rappe, A. K.; Goddard III, W. A. *J. Phys. Chem.* **1991**, *95*, 3358–3363.

[212] König, P.; Hoffmann, M.; Frauenheim, T.; Cui, Q. *J. Phys. Chem. B* **2005**, *109*, 9082.

[213] Zhang, Y.; Lee, T.; Yang, W. *J. Chem. Phys. B* **1999**, *110*, 46.

[214] Antes, I.; Thiel, W. *J. Phys. Chem. A* **1999**, *103*, 9290.

[215] Ferenczy, G. G.; Rivail, J. L.; Surjan, P. R.; Naray-Szabo, G. *J. Comput. Chem.* **1992**, *13*, 830.

[216] Thery, V.; Rinaldi, D.; Rivail, J. L.; Maigret, B.; Ferenczy, G. G. *J. Comput. Chem.* **1994**, *15*, 269.

[217] Philipp, D. M.; Friesner, R. A. *J. Comput. Chem.* **1999**, *20*, 1468.

[218] Murphy, R. B.; Philipp, D. M.; Friesner, R. A. *J. Comput. Chem.* **2000**, *21*, 1442.

[219] Gao, J.; Amara, P.; Alhambra, C.; Field, M. J. *J. Phys. Chem. A* **1998**, *102*, 4714.

[220] Steinbach, P. J.; R., B. B. *J. Comput. Chem.* **1994**, *15*, 667.

[221] Stote, R. H.; Karplus, M. *Proteins* **1995**, *23*, 12.

[222] Feller, S. E.; Pastor, R. W.; Rojnuckarin, A.; Bogusz, S.; Brooks, B. R. *J. Phys. Chem.* **1996**, *100*, 17011.

[223] Stote, R. H.; States, D. J.; Karplus, M. *J. Chim. Phys.* **1991**, *88*, 2419.

[224] Sagui, C.; Darden, T. A. *Annu. Rev. Biophys. Biomol. Struct.* **1999**, *28*, 155.

[225] Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089.

[226] Kuwajima, S.; Warshel, A. *J. Chem. Phys.* **1988**, *89*, 3751.

[227] Nam, K.; Gao, J.; York, D. M. *J. Chem. Theo. Comput.* **2005**, *1*, 2.

[228] Weber, W.; Hunenberger, P. H.; McCammon, J. A. *J. Phys. Chem. B* **2000**, *104*, 3668.

[229] Simonson, T.; Archontis, G.; Karplus, M. *J. Phys. Chem. B* **1997**, *101*, 8349.

[230] Sharp, K. A.; Honig, B. *Annu. Rev. Biophys. Chem.* **1990**, *19*, 301.

[231] Davis, M. E.; McCammon, J. A. *Chem. Rev.* **1990**, *90*, 509.

[232] Li, G.; Cui, Q. *J. Phys. Chem. B* **2004**, *108*, 3342.

[233] Birge, R. R. *Biochim. Biophys. Acta* **1990**, *1016*, 293.

[234] Okamura, M.; Paddock, M.; Graige, M.; Feher, G. *Biochimica et Biophysica Acta* **2000**, *1458*, 148.

[235] Kreuer, K.; Paddison, S. J.; Spohr, E.; Schuster, M. *Chem. Rev.* **2004**, *104*, 4637.

[236] Fersht, A. *Structure and Mechanism in Protein Science. Guide to Enzyme Catalysis and Protein Folding;* W. H. Freeman, 1999.

[237] Frank, R. A. W.; Titman, C. M.; Pratap, J. V.; Luisi, B. F.; Perham, R. N. *Science* **2004**, *306*, 872.

[238] Kohen, A.; Klinman, J. *Acc. Chem. Res.* **1998**, *31*, 397.

[239] Liang, Z.; Klinman, J. *Curr. Opin. Struct. Biol.* **2004**, *14*, 648.

[240] Hwang, J. K.; Warshel, A. *J. Am. Chem. Soc.* **1996**, *118*, 11745.

[241] Cui, Q.; Karplus, M. *J. Am. Chem. Soc.* **2002**, *124*, 3093.

[242] Kiefer, P. M.; Hynes, J. T. *J. Phys. Chem. A* **2003**, *107*, 9022.

[243] Hammes-Schiffer, S. *Biochem.* **2002**, *41*, 13335.

[244] Sham, Y. Y.; Muegge, I.; Warshel, A. *Proteins: Struct., Funct., Genet.* **1999**, *36*, 484.

[245] Nagle, J. F.; Morowitz, H. J. *Proc. Nat. Acad. Sci.* **1978**, *75*, 298.

[246] Nagle, J. F.; Mille, M. *J. Chem. Phys.* **1981**, *74*, 1367.

[247] Åqvist, J.; Warshel, A. *J. Mol. Biol.* **1992**, *224*, 7.

[248] Braun-Sand, S.; Strajbl, M.; Warshel, A. *Biophys. J.* **2004**, *87*, 2221.

[249] Schutz, C. N.; Warshel, A. *J. Phys. Chem. B* **2004**, *108*, 2066.

[250] Xu, D.; Riccardi, D.; Ghosh, N.; Elstner, M.; Guo, H.; Cui, Q. *J. Am. Chem. Soc. Comm. Ed.* **2005**, *Submitted*.

[251] Toba, S.; Colombo, G.; Merz, K. M. J. *J. Am. Chem. Soc.* **1999**, *121*, 2290.

[252] Lu, D.; Voth, G. A. *Proteins: Struct., Funct., and Genet.* **1998**, *33*, 119.

[253] Pomès, R.; Roux, B. *Biophys. J.* **1998**, *75*, 33.

[254] Bondar, A.; Fischer, S.; Smith, J.; Elstner, M.; Suhai, S. *J. Am. Chem. Soc.* **2004**, *126*, 14668.

[255] Bondar, A.; Elstner, M.; Suhai, S.; Smith, J. C.; Fischer, S. *Structure* **2004**, *12*, 1281.

[256] Wu, Y.; Voth, G. A. *Biophys. J.* **2003**, *85*, 864.

[257] Popovic, D. M.; Stuchebrukhov, A. A. *J. Am. Chem. Soc.* **2004**, *126*, 1858.

[258] Cukier, R. I. *Biochim. Biophys. Acta.* **2005**, *1706*, 134.

[259] Olkhova, E.; Huter, M. C.; Lill, M. A.; Helms, V.; Michel, H. *Biophys. J.* **2004**, *86*, 1873.

[260] Wikstrom, M.; Verkhovsky, M. I.; Hummer, G. *Biochim. Biophys. Acata* **2003**, *1604*, 61.

[261] Wikstrom, M. *Curr. Opin. Struct. Biol.* **1998**, *8*, 480.

[262] König, P. H. *Modeling Long-Range Proton Transfer: New Developments and Application to the Photosynthetic Bacterial Reaction Center.* PhD thesis, Universität Paderborn, **2005**.

[263] Smondyrev, A. M.; Voth, G. A. *Biophys. J.* **2002**, *83*, 1987.

[264] Bolhuis, P. G.; Chandler, D.; Dellago, C.; Geissler, P. L. *Annu. Rev. Phys. Chem.* **2002**, *53*, 291.

[265] Florian, J.; Warshel, A. *J. Phys. Chem. B* **1997**, *101*, 5583.

[266] Shurki, A.; Warshel, A. *Adv. Prot. Chem.* **2003**, *66*, 249.

[267] Mo, Y. R.; Gao, J. *J. Comput. Chem.* **2000**, *21*, 1458.

[268] Eigen, M. *Angew. Chem.* **1963**, *75*, 489–508.

[269] Zundel, G.; Metzger, H. *Z. Phys. Chemie* **1968**, *58*, 225.

[270] Yarnell, A. *Chem. Eng. News.* **2004**, *81*, 42.

[271] Im, W.; Beglov, D.; Roux, B. *Comput. Phys. Comm.* **1998**, *111*, 59.

[272] Woo, H. J.; Dinner, A. R.; Roux, B. *J. Chem. Phys.* **2004**, *121*, 6392.

[273] Torrie, G. M.; Valleau, J. P. *J. Comput. Phys.* **1977**, *23*, 187.

[274] Warshel, A.; Sussman, F.; King, G. *Biochemistry* **1986**, *25*, 8368.

[275] Maurel, C.; Reizer, J.; Schroeder, J. L.; Chrispeels, M. J.; Saier, M. H. *J. Biol. Chem.* **1994**, *269*, 11869.

[276] Park, J. H.; Saier, M. H. *J. Membr. Biol.* **1996**, *153*, 171.

[277] Saparov, S. M.; Tsunoda, S. P.; Pohl, P. *Biol. Cell.* **2005**, *97*, 545.

[278] Decad, G. M.; Nikaido, H. *J. Bacteriol.* **1976**, *128*, 325.

[279] Fischer, A. **1903**.

[280] de Groot, B. L.; Engel, A.; Grubmüller, H. *FEBS Lett.* **2001**, *504*, 206.

[281] Hiroaki, Y.; Tani, K.; Kamegawa, A.; Gyobu, N.; Nishikawa, K.; Suzuki, H.; Walz, T.; Sasaki, S.; Mitsuoka, K.; Kimura, K.; Mizoguchi, A.; Fujiyoshi, Y. *J. Mol. Biol.* **2005**, *355*, 628.

[282] Tornroth-Horsefield, S.; Wang, Y.; Hedfalk, K.; Johanson, U.; Karlsson, M.; Tajkhorshid, E.; Neutze, R.; Kjellbom, P. *Nature (London)* **2006**, *439*, 688.

[283] Lee, J. K.; Kozono, D.; Remis, J.; Kitagawa, Y.; Agre, P.; Stroud, R. M. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 18932.

[284] Heymann, J. B.; Engel, A. *J. Mol. Biol.* **2000**, *295*, 1039.

[285] Nollert, P.; Harries, W.; Fu, D.; Miercke, L. J. W.; Stroud, R. M. *FEBS Letters* **2001**, *504*, 112.

[286] Wistow, G. J.; Pisano, M. M.; Chepelinsky, A. B. *Tends Biochem. Sci.* **1991**, *16*, 170.

[287] Pao, G. M.; Wu, L. F.; Johnson, J. D.; Hofte, H.; Chrispeels, M. J.; Sweet, G.; Sandal, N. N.; Saier, M. H. *Mol. Microbiol.* **1991**, *5*, 33.

[288] Dutzler, R.; Campbell, E. B.; Cadene, M.; Chait, B. T.; MacKinnon, R. *Nature (London)* **2002**, *415*, 287.

[289] Abramson, J.; Smirnova, I.; Kasho, V.; Verner, G.; Kaback, H. R.; Iwata, S. *Science* **2003**, *301*, 610.

[290] Stroud, R. M.; Savage, D.; Miercke, L. J. W.; Khademi, S.; Lee, J. K.; Harries, W. *FEBS Letters* **2003**, *555*, 79.

[291] Wang, Y.; Schulten, K.; Tajkhorshid, E. *Structure* **2005**, *13*, 1107.

[292] Jensen, M. O.; Tajkhorshid, E.; Schulten, K. *Structure* **2001**, *9*, 1083.

[293] Eisenberg, B. *Biophys. J.* **2003**, *85*, 3427.

[294] de Groot, B. L.; Grubmueller, H. *Curr. Opin. Struct. Biol.* **2005**, *15*, 176.

[295] Sui, H.; Han, B. G.; Lee, J. K.; Walian, P.; Jap, B. K. *Nature (London)* **2001**, *414*, 872.

[296] Zhu, F.; Tajkhorshid, E.; Schulten, K. *FEBS Lett.* **2001**, *504*, 212.

[297] Jensen, M. J.; Röthlisberger, U.; Rovira, C. *Biophys. J.* **2005**, *89*, 1744.

[298] Nina, M.; Beglov, D.; Roux, B. *J. Phys. Chem. B* **1997**, *101*, 5239.

[299] Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684.

[300] Smart, O. S.; Goodfellow, J. M.; Wallace, B. A. *Biophys. J.* **2003**, *65*, 2455.

[301] Bash, P. A.; Field, M. J.; Davenport, R. C.; Petsko, G. A.; Ringe, D.; Karplus, M. *Biochemistry* **1991**, *30*, 5826–32.

[302] Roux, B.; Karplus, M. *Biophys. J.* **1991**, *59*, 961–81.

[303] König, P. H.; Cui, Q.; et al.. manuscript in preparation.

[304] Roux, B.; Karplus, M. *Annu. Rev. Biophys. Biomol. Struct.* **1994**, *23*, 731–761.

[305] Hummer, G.; Rasaiah, J. C.; Noworyta, J. P. *Nature* **2001**, *414*, 188.

[306] Beckstein, O.; Sansom, M. S. P. *Proc. Natl. Acad. Sci., USA* **2003**, *100*, 7063.

[307] Allen, R.; Melchionna, S.; Hansen, J. P. *Phys. Rev. Lett.* **2002**, *89*, 175502–1.

[308] Tepper, H. L.; Voth, G. A. *Biophys. J.* **2005**, *88*, 3095.

[309] Marrink, S.; Jähnig, F.; Berendsen, H. J. C. *Biophys. J.* **1996**, *71*, 632.

[310] Lu, D.; Rotkin, S. V.; Ravaioli, U.; Schulten, K. *Nano Lett.* **2004**, *4*, 2383.

[311] Sengupta, D.; Behera, R. N.; Smith, J. C.; Ullmann, G. M. *Structure* **2005**, *13*, 849.

[312] Åqvist, J.; Luecke, H.; Quiocho, F. A.; Warshel, A. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 2026.

[313] Haupts, U.; Tittor, J.; Oesterhelt, D. *Annu. Rev. Biophys. Biomol. Struct.* **1999**, *28*, 367.

[314] Lanyi, J. *J. Phys. Chem. B* **2000**, *104*, 11441.

[315] Lanyi, J. *Annu. Rev. Physiol.* **2004**, *66*, 144–167.

[316] Oesterhelt, D.; Stoeckenius, W. *Proc. Natl. Acad. Sci. U.S.A.* **1973**, *70*, 2853.

[317] Spudich, J. L.; Bogomolni, R. A. *Annu. Rev. Biophys. Biophys. Chem.* **1988**, *17*, 193.

[318] Hoff, W. D.; Jung, K.; Spudich, J. L. *Annu. Rev. Biophys. Biomol. Struct.* **1997**, *26*, 223.

[319] Lanyi, J. K. *Annu. Rev. Biophys. Biophys. Chem.* **1986**, *15*, 11.

[320] Luecke, H.; Schobert, B.; Lanyi, J.; Spudich, E. N.; Spudich, J. L. *Science* **2001**, *293*, 1499–1503.

[321] Kamo, N.; Shimono, K.; Iwamoto, M.; Sudo, Y. *Biochemistry (Moscow)* **2001**, *66*, 1277.

[322] Shimono, K.; Kitami, M.; Iwamoto, M.; Kamo, N. *Biophys. Chem.* **2000**, *87*, 225.

[323] Shimono, K.; Furutani, Y.; Kandori, H.; Kamo, N. *Biochemistry* **2002**, *41*, 6504–6509.

[324] Shimono, K.; Hayashi, T.; Ikeura, Y.; Sudo, Y.; Iwamoto, M.; Kamo, N. *J. Biol. Chem.* **2003**, *278*, 23882.

[325] Shimono, K.; Iwamoto, M.; Sumi, M.; Kamo, N. *Photochem. Photobiol.* **2000**, *72*, 141.

[326] Grigorieff, N.; Ceska, T. A.; Downing, K. H.; Baldwin, J. M.; Henderson, R. *J. Mol. Biol.* **1996**, *259*, 393.

[327] Kimura, Y.; Vassylyev, D. G.; Miyazawa, A.; Kiera, A.; Matsushima, M. *Nature (London)* **1997**, *389*, 206.

[328] Landau, E. M.; Rosenbusch, J. P. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 14532.

[329] Schobert, B.; Cupp-Vickery, J.; Hornak, V.; Smith, S.; Lanyi, J. *J. Mol. Biol.* **2002**, *321*, 715.

[330] Belrhali, H.; Nollert, P.; Royant, A.; Menzel, C.; Rosenbusch, J.; Landau, E.; Pebay-Peyroula, E. *Structure* **1999**, *7*, 909.

[331] Luecke, H.; Richter, H. T.; Lanyi, J. K. *Science* **1998**, *280*, 1934.

[332] Luecke, H.; Schobert, B.; Richter, H. T.; Cartailler, J. P.; Lanyi, J. K. *J. Mol. Biol.* **1999**, *291*, 899–911.

[333] Essen, L.; Siegert, R.; Lehmann, W. D.; Oesterhelt, D. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 11673.

[334] Smith, S. O.; Myers, A. B.; Pardoen, J. A.; Winkel, C.; Mulder, P. P. J.; Lugtenburg, J.; Mathies, R. *Proc. Natl. Acad. Sci. U.S.A.* **1984**, *81*, 2055.

[335] Lozier, R. H.; Niederberger, W.; Ottolenghi, M.; Sivorinowski, G.; Stockenius, W. In *Energetics and structure of halophilic microorganisms.*; Elsevier/North-Holland Biomedical Press, 1979; pages 123–141.

[336] Sperling, W.; Rafferty, C. N.; Kohl, K.; Dencher, N. A. *FEBS Lett.* **1979**, *97*, 129.

[337] Fahmy, K.; Siebert, F.; F., G. M.; Tavan, P. *J. Mol. Struct.* **1989**, *214*, 257.

[338] Huang, J. Y.; Lewis, A. *Biophys. J.* **1989**, *55*, 835.

[339] Lin, S. W.; Mathies, R. A. *Biophys. J.* **1989**, *56*, 635.

[340] Rothschild, K. J.; Gray, D.; Mogi, T.; Marti, T.; Braiman, M. S.; Stern, L. J.; Khorana, H. G. *Biochemistry* **1989**, *28*, 7052.

[341] Riesle, J.; Oesterheld, D.; Dechner, N. A.; Heberle, J. *Biochemistry* **1996**, *35*, 6635.

[342] Otto, H.; Marti, T.; Holz, M.; Mogi, T.; Lindau, M.; Khorana, H. G.; Heyn, M. P. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 9228.

[343] Gerwert, K.; Hess, B.; Soppa, J.; Oesterhelt, D. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 4943.

[344] Coutre, J. L. Tittor, J.; Oesterhelt, D.; Gerwert, K. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 4962.

[345] Roux, B.; Nina, M.; Pomès, R.; Smith, J. C. *Biophys. J.* **1996**, *71*, 670.

[346] Humphrey, W.; Logunov, I.; Schulten, K.; Sheves, M. *Biochemistry* **1994**, *33*, 3668.

[347] Lanyi, J. K. *Biochim. Biophys. Acta* **1993**, *1183*, 241.

[348] Lanyi, J. K. *Nature (London)* **1995**, *375*, 461.

[349] Lanyi, J. K. *J. Struct. Biol.* **1998**, *124*, 164.

[350] Hage, W.; Kim, M.; Frei, H.; Mathies, R. A. *J. Phys. Chem.* **1996**, *100*, 16026.

[351] Bullough, P. A.; Henderson, R. *J. Mol. Biol.* **1999**, *286*, 1663.

[352] Doig, S. J.; Reid, P. J.; Mathies, R. A. *J. Phys. Chem.* **1991**, *95*, 6372.

[353] Brack, T. L.; Atkinson, G. H. *J. Mol. Struct.* **1989**, *214*, 289.

[354] Edman, K.; Nollert, P.; Royant, A.; Belrhali, H.; Pebay-Peyroula, E.; Hajdu, J.; Neutze, R.; Landau, E. M. *Nature (London)* **1999**, *401*, 822.

[355] Matsui, Y.; Sakai, K.; Murakami, M.; Shiro, Y.; Adachi, S.; H., O.; Kouyama, T. *J. Mol. Biol.* **2002**, *324*, 469.

[356] Gerwert, K.; Souvignier, G.; Hess, B. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 9774.

[357] Drachev, L. A.; Kaulen, A. D.; Skulachev, V. P. *FEBS Lett.* **1984**, *178*, 331.

[358] Grzesiek, S.; Dencher, N. A. *FEBS Lett.* **1986**, *208*, 337.

[359] Heberle, J.; Dencher, N. A. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 5995.

[360] Brown, L. S.; Sasaki, J.; Kandori, H.; Maeda, A.; Needleman, R.; Lanyi, J. K. *J. Biol. Chem.* **1995**, *270*, 27122 – 27126.

[361] Balashov, S. P.; Imasheva, E. S.; Ebrey, T. G.; Chen, N.; Menick, D. R.; Crouch, R. K. *Biochemistry* **1997**, *36*, 8671.

[362] Rammelsberg, R.; Huhn, G.; Lubben, M.; Gerwert, K. *Biochemistry* **1998**, *37*, 5001.

[363] Spassov, V. Z.; Luecke, H.; Gerwert, K.; Bashford, D. *J. Mol. Biol.* **2001**, *312*, 203.

[364] Subramaniam, S.; Lindahl, M.; Bullough, P.; Faruqi, A. R.; Tittor, J.; Oesterhelt, D.; Brown, L.; J., L.; Henderson, R. *J. Mol. Biol.* **1999**, *287*, 145.

[365] Ormos, P. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 473.

[366] Varo, G.; Lanyi, J. K. *Biochemistry* **1991**, *30*, 5061.

[367] Ames, J. B.; Methies, R. A. *Biochemistry* **1990**, *29*, 7181.

[368] Gai, F.; C., H. K.; McDonald, J. C.; Anfinrud, P. A. *Science* **1998**, *279*, 1886.

[369] Weidlich, O.; Schalt, B.; Friedman, N.; Sheves, M.; Lanyi, J. K.; Brown, L. S.; Siebert, F. *Biochemistry* **1996**, *35*, 10807.

[370] Zscherp, C.; Heberle, J. *J. Phys. Chem.* **1997**, *101*, 10542.

[371] Bivin, D. B.; Stoeckenius, W. *J. Gen. Microbiol.* **1986**, *132*, 2167.

[372] Hirayama, J.; Imamoto, Y.; Shichida, Y.; Kamo, N.; Tomioka, H.; Yoshizawa, T. *Biochemistry* **1992**, *31*, 2093.

[373] Takahshi, T.; Yan, B.; Mazur, P.; Derguini, F.; Nakanishi, K.; Spudich, J. L. *Biochemistry* **1990**, *29*, 8467.

[374] Zhang, X.; Zhu, J.; Spudich, J. L. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 857.

[375] Wegener, A.; Klare, J. P.; Engelhard, M.; Steinhoff, H. *EMBO J.* **2001**, *20*, 5312.

[376] Gordeliy, V.; Labahn, J.; Moukhametzianov, R.; Efremov, R.; Granzin, J.; Schlesinger, R.; Bueldt, G.; Savopol, T.; Scheidig, A.; Klare, J. andEngelhard, M. *Nature (London)* **2002**, *419*, 484.

[377] Moukhametzianov, R.; Klare, J. P.; Efremov, R.; Baeken, C.; Göppner, A.; Labahn, J.; Engelhard, M.; Büldt, G.; Gordeliy, V. I. *Nature (London)* **2006**, *440*, 115.

[378] Royant, A.; Nollert, P.; Edmann, K.; Neutze, R.; Landau, E. M.; Pebay-Peyroula, E.; Navarro, J. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 10131–10136.

[379] Iwamoto, M.; Shimono, K.; Sumi, M.; Kamo, N. *Biophys. Chem.* **1999**, *79*, 187.

[380] Schmies, G.; Lüttenberg, B.; Chizhov, I.; Engelhard, M.; Becker, A.; Bamberg, E. *Biophys. J.* **2000**, *78*, 967.

[381] Sudo, Y.; Iwamoto, M.; Shimono, K.; Sumi, M.; Kamo, N. *Biophys. J.* **2001**, *80*, 916.

[382] Schmies, G.; Engelhard, M.; Wood, P. G.; Nagel, G.; Bamberg, E. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 1555.

[383] Sasaki, J.; Spudich, J. L. *Biophys. J.* **1999**, *77*, 2145.

[384] Seidel, R.; Scharf, B.; Gautel, M.; Kleine, K.; Oesterhelt, D.; Engelhard, M. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 3036.

[385] Wada, A.; Akai, A.; Goshima, T.; Takahashi, T.; Ito, M. *Bioorg. Med. Chem. Lett.* **1998**, *8*, 1365.

[386] Hirayma, J.; Kamo, N.; Imamoto, Y.; Y., S.; Yoshizawa, T. *FEBS Lett.* **1995**, *364*, 168.

[387] Sasaki, J.; Spudich, J. L. *Biochim. Biophys. Acta* **2000**, *1460*, 230.

[388] Yan, B.; Takahashi, T.; Johnson, R.; Spudich, J. L. *Biochemistry* **1991**, *30*, 10686.

[389] Wegener, A.; Chizhov, I.; M., E.; Steinhoff, H. *J. Mol. Biol.* **2000**, *301*, 881.

[390] Klare, J. P.; Gordeliy, V. I.; Labahn, J.; Büldt, G.; Steinhoff, H.; Engelhard, M. *FEBS Lett.* **2004**, *564*, 219.

[391] Van der Steen, R.; Biesheuvel, B. L.; Lugtenburg, J. *J. Am. Chem. Soc.* **1986**, *108*, 6410.

[392] Harbison, G. S.; Mulder, P. P. J.; Pardoen, J. A. *J. Am. Chem. Soc.* **1985**, *107*, 4810.

[393] Wada, M.; Sakutai, M.; Inoue, Y.; Tamura, Y.; Watanabe, Y. *J. Am. Chem. Soc.* **1994**, *116*, 1537.

[394] Kakitani, H.; Kakitani, T.; Rodman, H.; Honig, B. *Photochem. Photobiol.* **1985**, *41*, 471.

[395] Blatz, P. E.; Liebman, P. A. *Exp. Eye Res.* **1973**, *17*, 573.

[396] Irving, C. S.; Byers, G. W.; Leermake, P. A. *J. Am. Chem. Soc.* **1969**, *91*, 2141.

[397] Irving, C. S.; Byers, G. W.; Leermake, P. A. *Biochemistry* **1970**, *9*, 858.

[398] Beppu, Y.; Kakitani, T. *Photochem. Photobiol.* **1994**, *59*, 660.

[399] Polland, H. J.; Franz, M. A.; Zinth, W.; Kaiser, W.; Oesterhelt, D. *Biocheim. Biophys. Acta* **1986**, *851*, 407.

[400] Birge, R. R.; Murray, L. M.; Pierce, B. M.; Akita, H.; Balogh-Nair, V.; Findsen, L. A.; Nakanishi, K. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 4117–4121.

[401] Aharoni, A.; Khatchatouriants, A.; Manevitch, A.; Lewis, A.; Sheves, M. *J. Phys. Chem. B* **2003**, *107*, 6221.

[402] Kandori, H.; Shimono, K.; Sudo, Y.; Iwamoto, M.; Shichida, Y.; Kamo, N. *Biochemistry* **2001**, *40*, 9238–9246.

[403] Greenhalgh, D. A.; Farrens, D. L.; Subramaniam, S.; Khorana, H. G. *J. Biol. Chem.* **1993**, *268*, 20305.

[404] Mogi, T.; Marti, T.; Khorana, H. G. *J. Biol. Chem.* **1989**, *264*, 14197.

[405] Hackett, N. R.; Stern, L. J.; Chao, B. H.; Kronis, K. A.; Khorana, H. G. *J. Biol. Chem.* **1987**, *262*, 9277.

[406] Ahl, P. L.; Stern, L. J.; Düring, D.; Mogi, T.; Khorana, H. G.; Rothschild, K. J. *J. Biol. Chem.* **1988**, *263*, 13594.

[407] Shimono, K.; Furutani, Y.; Kamo, N.; Kandori, H. *Biochemistry* **2003**, *42*, 7801–7806.

[408] Balashov, S. P.; Govindjee, R.; Kono, M.; Imasheva, E.; Lukashev, E.; Ebrey, T. G.; Crouch, R. K.; Menick, D. R.; Feng, Y. *Biochemistry* **1993**, *32*, 10331.

[409] Ikeura, Y.; Shimono, K.; Iwamoto, M.; Sudo, Y.; Kamo, N. *Photochem. Photobiol.* **2003**, *77*, 96.

[410] Wanko, M.; Hoffmann, M.; Strodel, P.; Koslowski, A.; Thiel, W.; Neese, F.; Frauenheim, T.; Elstner, M. *J. Phys. Chem. B* **2005**, *109*, 3606–3615.

[411] Koslowski, A.; Beck, M. E.; Thiel, W. *J. Comput. Chem.* **2003**, *24*, 714–726.

[412] Neese, F. *J. Chem. Phys.* **2003**, *119*, 9428.

[413] Page, C. S.; Olivucci, M. *J. Comput. Chem.* **2003**, *24*, 298–309.

[414] Fantacci, S.; Migani, A.; Olivucci, M. *J. Phys. Chem. A* **2004**, *108*, 1208–1213.

[415] Zhou, H.; Tajkhorshid, E.; Frauenheim, T.; Suhai, S.; Elstner, M. *Chem. Phys.* **2002**, *277*, 91–103.

[416] Vreven, T.; Morokuma, K. *Theor. Chem. Acc.* **2003**, *109*, 125.

[417] Hayashi, S.; Ohmine, I. *J. Phys. Chem. B* **2000**, *104*, 10678–10691.

[418] Neese, F. "ORCA – An ab initio, density functional and semiempirical program package, Version 2.3 - Revision 09, February 2004", Max Planck Institut für Strahlenchemie, Mülheim., **2004**.

[419] Schaefer, A.; Horn, H.; Ahlrichs, R. *J. Chem. Phys.* **1992**, *97*, 2571.

[420] Strodel, P.; Tavan, P. *J. Chem. Phys.* **2002**, *117*(10), 4677.

[421] Sasaki, J.; Lany, J. K.; Needleman, R.; Yoshizawa, T.; Maeda, A. *Biochemistry* **1994**, *33*, 3178.

[422] Dinner, A. R.; Lopez, X.; Karplus, M. *Theor. Chem. Acc.* **2003**, *109*, 118.

[423] Nose, S. *J. Chem. Phys.* **1984**, *81*, 551.

[424] Hoover, W. *Phys. Rev. A* **1985**, *31*, 1695.

[425] Nonella, M.; Mathias, G.; Tavan, P. *J. Phys. Chem. A* **2003**, *107*, 8638.

[426] Kandori, H.; Furutani, Y.; Shimono, K.; Shichida, Y.; Kamo, N. *Biochemistry* **2001**, *40*, 15693–15698.

[427] Birge, R. R.; Zhang, C. *J. Chem. Phys.* **1990**, *92*, 7178.

[428] Mathies, R.; Stryer, L. *Proc. Natl. Acad. Sci. U.S.A.* **1976**, *73*, 2169.

[429] González-Luque, R.; Garavelli, M.; Bernardi, F.; Merchán, M.; Robb, M. A.; Olivucci, M. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 9379.

[430] Cembran, A.; Bernardi, F.; Olivucci, M.; Garavelli, M. *J. Am. Chem. Soc.* **2003**, *125*, 12509.

[431] Hoffmann, M.; Wanko, M.; Strodel, P.; König, P. H.; Frauenheim, T.; Schulten, K.; Thiel, W.; Elster, M. *J. Am. Chem. Soc.* **2006**, *in print*.

[432] Hufen, J.; Sugihara, M.; Buss, V. *J. Phys. Chem.* **2004**, *108*, 20419.

[433] Böttcher, C. J. F. *Theory of Electric Polarization.*, Vol. 1; Elsevier Scientific Publ. Co.: Amsterdam, 2 ed., 1973.

[434] Hayashi, S.; Tajkhorshid, E.; Kandori, H.; Schulten, K. *J. Am. Chem. Soc.* **2004**, *126*, 10516.

[435] Kochendoerfer, G.; Lin, S. W.; Sakmar, T. P.; Mathies, R. A. *Trends Biochem. Sci.* **1999**, *24*, 300.

[436] Baasov, T.; Sheves, M. *J. Am. Chem. Soc.* **1985**, *107*, 7524.

[437] Heyde, M. E.; Gill, D.; Kilponen, R. G.; Rimai, L. *J. Am. Chem. Soc.* **1979**, *93*, 6776.

[438] Zhang, Y.; Kua, J.; McCammon, J. A. *J. Phys. Chem. B* **2003**, *107*, 4459.

[439] Lin, S. W.; Groesbeek, M.; van der Hoef, I.; Verdegem, P.; Lugtenburg, J.; Mathies, R. A. *J. Phys. Chem. B* **1998**, *102*, 2787.

[440] Hoffmann, M.; König, P. H.; et al.. manuscript in preparation.

[441] Mostaghim, S.; Hoffmann, M.; König, P. H.; Frauenheim, T.; Teich, J. *Proceedings of the 2004 IEEE Congress on Evolutionary Computation* **2004**, page 212.

[442] Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graph.* **1996**, *14*, 33.

[443] Kraulis, P. J. *J. Appl. Cryst.* **1991**, *24*, 946.

[444] Merrit, E. A.; Bacon, D. J. *Meth. Enzymol.* **1997**, *277*, 505.

[445] Smart, O. S.; Goodfellow, J. M.; Wallace, B. A. *Biophys. J.* **2003**, *65*, 2455.

Own Publications

1. *Color tuning in Rhodopsins: the mechanism for the spectral shift between Bacteriorhodopsin and Sensory Rhodopsin II.*,
M. Hoffmann, M. Wanko, P. Strodel, P.H. König, Th. Frauenheim, K. Schulten, W. Thiel, M. Elster,
Journal of the American Chemical Society, 2006, 128, 10808.

2. *Calculating Absorption Shifts for Retinal Proteins: Computational Challenges.*,
M. Wanko, M. Hoffmann, P. Strodel, A. Koslowski, W. Thiel, F. Neese, Th. Frauenheim, M. Elstner,
Journal of Physical Chemistry B, 2005, 109, 3606.

3. *A critical evaluation of different QM/MM frontier treatments using SCC-DFTB as the QM method*,
P. H. König*, M. Hoffmann*, Th. Frauenheim and Q. Cui,
Journal of Physical Chemistry B, 2005, 109, 9082.

4. *Molecular Force Field Parametrization using Multi-Objective Evolutionary Algorithms*,
S. Mostaghim*, M. Hoffmann*, P. H. König*, Th. Frauenheim, J. Teich,
Proceedings of the 2004 IEEE Congress on Evolutionary Computation (2004), 212. DOI: 10.1109/CEC.2004.1330859.

5. *Towards theoretical analysis of proton transfer kinetics in biomolecular pumps*,
P.H. König, N. Ghosh, M. Hoffmann, M. Elstner, E. Tajkhorshid, Th. Frauenheim, Q. Cui,
Journal of Physical Chemistry A, 2006, 110, 548.

6. *Computational Photochemistry of Retinal Proteins*,
M. Wanko, M. Hoffmann, Th. Frauenheim, M. Elstner,
Journal of Computer Aided Molecular Design, in print.

* These authors contributed in a similar matter to publication.

Pieces of this thesis were published in (1) and (5).

Acknowledgements

First of all, I am most grateful to Prof. Dr. Thomas Frauenheim for accepting me into his group, his support during the last years and the freedom to expand my research activities according to my own interest. Further, I wish to thank Prof. Dr. Thomas Frauenheim and Prof. Dr. Marcus Elstner for providing me the opportunity to work on this most interesting topic.

Financially the research was supported by the Deutsche Forschungsgemeinschaft through the Forschergruppe 490 "Molecular Mechanisms of Retinal Protein Action".

I would like to express my gratitude to Prof. Dr. Klaus Schulten and Dr. Emad Tajkhorshid (Beckmann Institute,University of Illinois at Urbana-Champaign) for the fruitful and ongoing collaboration and their support in the color tuning [431] and aquaporin [130, 440] project. I also wish to thank them for making possible my visits to Urbana-Champaign and the sincere welcome in their group at Beckmann Institute.

Prof. Dr. Qiang Cui (University of Wisconsin, Madison) provided inspiring ideas and the implementation of the GSBP approach, without which the aquaporin project would not have been possible. The work on the reaction coordinate [130], the aquaporin project [130] as well as projects not mentioned in this thesis [212] result from this stimulating collaboration. I also thank him for his support and the pleasant stay in Madison.

I am most grateful to Prof. Dr. Walther Thiel for providing the OM2/MRCI method, his interest in the color tuning project, and the most helpful support in the preparation of the publication [431].

Dr. Peter König has been a belting comrade for the last years. Our meeting emerged as momentous fortuity. An extraordinary cooperation evolved that breeded a multiplicity of projects [130, 212, 431, 440, 441] on the one hand because of scientific necessity but also because of spontaneous sometimes loony ideas. Moreover, I would like to thank him for the great time outside the lab.

I thank my colleagues Marius Wanko and Dr. Paul Strodel for the stimulating collaboration. The discussions we had and the astir exchange created an outstanding productive collaboration [410, 431].

For sharing his theoretical and practical experience with me and his daily company during lunch at a small restaurant in Urbana-Champaign, I would like to thank Dr. Ulrich Kleinekathöfer.

One of the project would have not been possible without Dr. Sanaz Mostaghim, who shared her expertise in multiobjective optimization. The idea to use her method to parameterize force fields led to an interesting collaboration crowned with a publication [441].

Special thanks go to Dr. Christof Köhler, Jan Knaup (Universität Bremen), Dr. Marc Amkreutz (IFAM Bremen), Edda Kloppmann, and Prof. Dr. Mathias Ullmann (Universität Bayreuth) for the excitatory scientific discussions.

Most computations were carried out on the Frauenheim group computer cluster, the joint cluster with the PC^2 (PLING), and at the regional computing center in Cologne

(RRZK). I would like to thank the IT-team of the theoretical physics for the most stable, perfectly maintained computer system providing incredible resources.

I also thank the secretaries of the group, Simone Lange and Astrid Canisius for the great support regarding all kinds of administrative issues.

In addition, I would like to thank all people responsible for the marvelous atmosphere in the lab and not to forget out of the lab. It started on a red-letter day with the dialogue: "Do not open the door! – Which door? – That door !!!" and continued with movie evenings, benchmark and barbecue parties. Thanks to Peter König, Marc Amkreutz, Petra Stammmeier, Christof Köhler, Marius Wanko, Jan Knaup, Pierro Altoé, David Heringer, Betty Szûcs, Simone Lange, and Simone Sanna for the great time.

My thanks to all those whom I forgot to mention. In this place, I also would like to thank the remaining coauthors Nilanjan Ghosh, Dr. Axel Koslowski, Prof. Dr. Frank Neese, and Prof. Dr. Jürgen Teich.

And finally, I would like to thank my parents, grandparents, and other family members. That the family is mentioned at the end of these acknowledgements does not seem to reflect its true importance. However in this way, it will be the last thing to be remembered. In this spirit, I thank my family for their support during the last years.

Colophon

Molecular graphics were prepared using VMD [442], Molscript [443], Raster3D [444] and the hole2 [445] program.