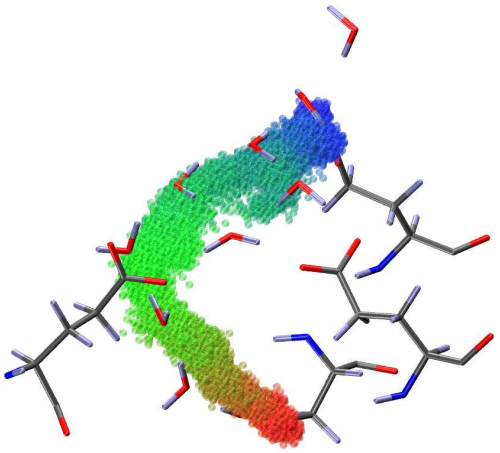
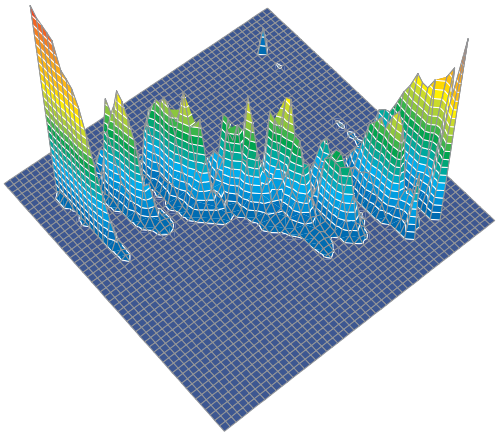
Modeling Long-Range Proton Transfer: New Developments and Application to the Photosynthetic Bacterial Reaction Center

Peter H. König





Modeling Long-Range Proton Transfer: New Developments and Application to the Photosynthetic Bacterial Reaction Center

Dissertation zur Erlangung des akademischen Grades

Doktor der Naturwissenschaften (Dr. rer. nat.)

vorgelegt dem

Department Physik der Fakultät für Naturwissenschaften

Universität Paderborn

Dipl. Chem. Peter H. König

Paderborn, 2005

Dem Department Physik der Fakultät für Naturwissenschaften als Dissertation vorgelegt.

Eingereicht am: 20.10.2005
Tag der mündlichen Prüfung: 8.12.2005

Promotionskommission

Vorsitzender	Prof. Dr. G. Wortmann
Erstgutachter	Prof. Dr. Th. Frauenheim
Zweitgutachter	Prof. Dr. G. Fels
Drittgutachter	Dr. Ch. Hoentzsch

Peter H. König, *Modeling Long-Range Proton Transfer: New Developments and Application to the Photosynthetic Bacterial Reaction Center*. PhD Thesis (in English), Department of Physics, Faculty of Science, University of Paderborn, Germany (2005).

Abstract

Proton transfer (PT) is essential for the working of metabolic processes in organisms from bacteria to humans. Short- and long-range PT are an integral part of reactions for instance in bioenergetics and enzymatic reactions in general. For the understanding of processes in cells on an atomistic level of detail, PT plays a central role. In this work, motivated by the goal of describing long-range PT, methodological developments for atomistic simulations were made, which were then applied to the bacterial reaction center (BRC).

First, for multiscale simulations, quantum mechanical/molecular mechanical (QM/MM) techniques were investigated for the CHARMM/SCC-DFTB framework. An issue of discussion in the recent literature is the description of electrostatics at the QM/MM boundary. In addition to previous suggestions for the QM/MM treatment of electrostatic interactions at the QM/MM boundary, a new approach, the divided frontier charge (DIV), was introduced and implemented. The performance of these schemes is evaluated based on properties including proton affinities, deprotonation energies, dipole moments, and energetics of proton transfer reactions. Similar to previous work, it is found that calculated proton affinities and deprotonation energies are very sensitive to the link atom scheme. Errors on the order of 15 to 20 kcal/mol are observed for less suitable models. Other schemes give better and, on average, mutually comparable results. For short-ranged proton transfer reactions, encouragingly, both activation barriers and reaction energies are fairly insensitive (within a typical range of 2-4 kcal/mol) to the link atom scheme due to error cancellation, and this was observed for both gas-phase and enzyme systems.

Another development presented here makes potential of mean force (PMF) calculations possible for long-range proton transfer. For this purpose, a set of collective reaction coordinates is proposed for characterizing the progress of long-range proton transfers. The modified center of excess charge (mCEC) overcomes some of the problems associated with previous suggestions. The mCEC can be used to define a new set of collective coordinates, which describe PT along highly nonlinear three-dimensional pathways, without specifying the mechanism or pathway *a priori*. The approaches of other authors were limited to linear pathways. Calculations on a realistic model of carbonic anhydrase demonstrate that adiabatic mapping using these collective coordinates gives reliable energetics and critical geometrical parameters similar to minimum energy path calculations. This suggests that the new coordinates can be effectively used as reaction coordinate in potential of mean force calculations for long-range PT in complex systems. Indeed, asside from the use in this project, the reaction coordinates presented here, are now in use by collaborators for a number of projects investigating PT.

As an application of the techniques developed here, the PMFs for proton transfer along a curved, $\sim$20 Å long pathway between Asp210(L) and Glu212(L) in the bacterial reaction center (BRC) were computed. The proton donor and acceptor are connected by a bifilar chain of eight water molecules and are 10 Å apart. Depending on the charge state of the quinones, the barrier for PT is 15-17 kcal/mol. This is slightly higher than what is expected from experiment. For the Q_B^- charge states, the transition state is earlier than the highest transition state found for the Q_B^0 charge states. Despite these phenomenological differences, the results for this model suggest that the sequence of events for proton transfer cannot solely be understood from direct electrostatic interaction with the quinones. Instead, the foundation may lay in an electrostatic domino effect, i.e. the rearrangement of salt-bridges between Q_A and the proton wire. In addition, the results show that the solvent structure changes significantly during the reaction, stressing again the importance of dynamical approaches to long-range proton transfer, which were made possible through the work presented here.

Peter H. König, *Modellierung von langreichweitigem Protonentransfer: Entwicklung neuer Methoden und Anwendung auf das photosynthetische bakterielle Reaktionszentrum*. Dissertation (in englischer Sprache), Department Physik, Fakultät für Naturwissenschaften, Universität Paderborn (2005).

Kurzfassung

Protonentransfer (PT) ist ein essentieller Baustein in metabolischen Prozessen aller Organismen, vom Bakterium zum Menschen. Kurz- und langreichweitiger PT spielt eine wichtige Rolle in der Bioenergetik und Reaktionen von Enzymen allgemein. Für ein Verständnis zellulärer Prozesse auf atomarer Ebene, ist die Beschreibung von Protonentransfer daher ein wichtiger Schritt. In dieser Arbeit werden eigene methodische Entwicklungen vorgestellt und auf das photosynthetische bakterielle Reaktionszentrum (BRC) angewendet.

Zunächst wurden quantenmechanische/molekularmechanische (QM/MM) Kopplungsmodelle für die SCC-DFTB/CHARMM Kombination untersucht. Ein offener Punkt der in der in den letzen Jahren diskutiert wurde, betrifft die Beschreibung der Elektrostatik an der QM/MM Grenze. Neben bereits bekannten Ansätzen, wurde hier ein eigener, neuer Vorschlag für die Beschreibung, *divided frontier charge* (DIV), untersucht. Die Modelle wurde an einer Vielzahl molekularer Eigenschaften geprüft, darunter Protonenaffinitäten (PA), Deprotonierungsenergien (DPE), Dipolmomente und die Energetik von PT-Reaktionen. Hier konnte bestätigt werden, daß die berechneten PA und DPA ausgesprochen empfindlich gegenüber der Wahl des Grenzatomschemas (*link atom*) sind. Für weniger geeignete Modelle ergeben sich Fehler im Bereich von 15-20 kcal/mol, wogegen andere Schemata bessere, qualitativ vergleichbare Ergebnisse zeigen. Erfeulicherweise sind Barrieren und Reaktionsenergien für kurzreichweitigen PT in Enzymen und in der Gasphase aufgrund von Fehlerkompensation deutlich unempfindlicher (2-4 kcal/mol).

Eine weitere hier präsentierte Entwicklung erlaubt die Berechnung des Potentials mittlerer Kraft (potential of mean force, PMF) für langreichweitigen PT. Dazu wurden neue Reaktionskoordinaten entwickelt. Mit der "modified center of excess charge" (mCEC) werden Probleme von vorherigen Vorschlägen zur Lokalisierung des Protonendefekts gelöst, die darüber hinaus auf lineare Pfade beschränkt waren. Die mCEC erlaubt es, einen neuen Satz kollektiver Koordinaten zu definieren, mit derren Hilfe PT in hochgradig gekrümmten, dreidimensionalen Pfaden beschrieben werden kann, ohne einen spezifischen Mechanismus *vorzugeben*. Simulationen an einem realistischem Modell von Carboanhydrase zeigen, daß *adiabatic mapping* mit diesen kollektiven Koordinaten zuverlässige Energien und Geometrien, in hervorragender Übereinstimmung mit minimalen Energiepfaden, liefert. Daher können diese neuen Koordinaten für die Berechnung von PMF für langreichweitigen PT eingesetzt werden. Neben der Verwendung in diesem Forschungsprojekt, befinden sich diese Methodiken bereits im Einsatz für Projekte von Kooperationspartnern.

Als Anwendung wurde das PMF für einen 20 Å langen PT Pfad im BRC berechnet. Der stark gekrümmte und verzweigte Pfad zwischen Asp210(L) und Glu212(L) besteht aus 8 Wassermolekülen. Dabei sind die beiden Endpunkte etwa 10 Å voneinander entfernt. Abhängig vom Ladungszustand der beiden Chinone beträgt die Barriere für diesen PT 15-17 kcal/mol. Dies ist im Vergleich zu experimentellen Ergebnissen geringfügig zu hoch. Für die Q_B^- Ladungszustände ist der Übergangszustand (ÜZ) etwas früher gelagert, als der höchste ÜZ für die Q_B^0 Ladungszustände. Trotz dieser phänomenologischen Unterschiede, zeigt dieses Modell, daß Abfolge der Schritte dieses Teils des Photozyklus nicht alleine anhand der direkten elektrostatischen Wechselwirkung mit den Chinonen verstanden werden kann. Stattdessen liegt der Unterschied möglicherweise in einem elektrostatischem Dominoeffekt begründet, d.h. einer Reorientierung der zahlreichen Salzbrücken zwischen Q_A und dem Protonentransferkanal. Eine weitere interessante Beobachtung ist, daß sich die Solvensstruktur während des Protonentransfers deutlich ändert. Dies hebt noch einmal die Bedeutung von dynamischen Ansätzen zur Beschreibung langreichweitigem PT hervor, die in durch die Entwicklungen dieser Arbeit ermöglicht werden.

Contents

List of Figures

List of Tables

Abreviations

au	atomic units
BO	born-Oppenheimer
CA	carboxylic anhydrase
BRC	bacterial reaction center
CEC	ceneter of excess charge
DFT	density functional theory
DFTB	density functional based tight binding
DHAP	dihydroxyacetone phosphate
DIV	divided frontier charges
DLA	double link atom
ET	electron transfer
EXGR	excluded groups
GSBP	generalized solvent boundary potential
mCEC	modified center of excess charge
MEP	minimum energy pathway
MGS	methyl glyoxal synthase
MM	molecular mechanics
NEB	nudged elastic band
PMF	potential of mean force
PT	proton transfer
QM	quantum mechanics/mechanical
RC	reaction center
rmsd	root mean square deviation
SLA	single link atom
SCC-DFTB	self-consistent charge density functional based tight binding
TIM	triosephosphate isomerase
QM/MM	quantum mechanical molecular mechanical coupling
WHAM	wheighted histograms analysis method

Chapter 1

Introduction

"Hieraus folgt aber, dass vollständige theoretische Erkenntnis,
in der wir die Verknüpfund der Thatsachen unter Gesetzen
aus diesen erklären, nur durch Mathematik
und nur so weit diese anwendbar ist, möglich wird."

M.J. Schleiden, 1849, in
"**Grundzüge der wissenschaftlichen Botanik nach einer methodologischen Einleitung als Anleitung zum Studium der Pflanze**"

1.1 Photosynthesis

Light takes about eight minutes to cover the distance between our sun and earth travelling at an unsurpassable 300000 km/s. This light is used to power a machinery which supplies more than 90 % of the energy consumed in the biosphere: photosynthesis. Like the light rays powering it, photosynthesis is an amazing marvel.

Since early phenomenological descriptions of photosynthesis more than 200 years ago, for instance by Priestley and Ingen-Housz [1], scientists of various disciplines have been working on uncovering the details.

Photosynthesis, in plants, algae and cyanobacteria, is the process, which is responsible for the light driven conversion of water and carbon dioxide to carbon hydrates and oxygen:

$$H_2O + CO_2 \xrightarrow{h\nu} (CHOH)_n + O_2$$

This phenomenological equation naturally hides the complexity of the processes occurring in photosynthesis. Based on experiments by Hill, Ruben and Camen in the first half of the twentieth century, it is known that the process of conversion of water to oxygen is only loosely connected to the process of carbon fixation: The overall process can be broken down into a light reaction and a light-independent, dark reaction.

The *light reaction* involves the absorption of light and the oxidation of water to oxygen. The site of this reaction are two distinct protein complexes. In plants, these are

light reaction

embedded in the thylakoid membranes of the chloroplasts, the photosynthetic reaction centers I and II (RC I, RC II). In the RCs, the freed hydrogen nuclei (protons) and electrons in the process of water oxidation are *pumped* electrochemically using a series of fine tuned and coupled electron- and proton-transfer processes. The protons are pumped across the membrane establishing a proton gradient. This proton gradient drives a *turbine*, the enzyme ATP-synthase, which synthesizes ATP (adenosine triphosphate), the universal energy currency of the cell. The electrons are stored in a high-energy chemical compound, a reduction agent, NADPH (nicotineamide adenine dinucleotide phosphate), which is required for numerous chemical metabolic processes.

dark reaction

The *dark reaction* uses both ATP and NADPH for synthesizing carbohydrates from absorbed carbon dioxide and water. The carbon fixation is accomplished by Rubisco (D-ribulose 1,5-bisphosphate carboxylase/oxygenase) an enzyme binding carbon dioxide to a sugar molecule [2], which is transformed in a series of reactions to other carbohydrates (Calvin cycle) [3].

bacterial reaction centers (BRC)

Despite the seemingly "inherent complexity´', evidence can be found that photosynthesis evolved gradually: Some photosynthetic bacteria contain only one reaction center, the *bacterial reaction centers (BRC)* [1]. These BRCs are very similar to either the RC I or RC II in sequence (i.e. the order of amino acids), structure and function [4, 5]. These bacteria are anoxygenic, i.e., non-oxygen producing, lacking the capability of water oxidation, and do not produce NADH (nicotineamide adenine dinucleotide, corresponding to NADPH in plants) in the light reaction.

Lacking the production of NADH in the light reaction, anoxygenic bacteria use other metabolic pathways to produce NADH, i.e. through the oxidation of organic molecules, hydrogen or sulfur compounds.

purple bacteria

One commonality to the diversity of all these different varieties of approaches to photosynthesis are the RCs, respectively the BRCs. The best characterized and understood RCs are the BRCs from *purple bacteria*, namely the species Rhodopseudomonas viridis, Rhodobacter sphaeroides. These BRC are similar to the RC II in plants. They have been used for studying principal aspects of photosynthesis, i.e., the conversion of light energy into chemical energy utilizable to power metabolical processes by using proton and electron transfer and pumping. In the following, the treatise will focus on the BRC in these species.

In the mid 1980's, Michel, Huber and Deisenhofer crystallized the BRC from *Rhodopseudomonas viridis* and elucidated its three dimensional atomic structure by X-Ray diffraction [6][2]. This was the first membrane protein to be crystallized and also the first membrane protein whose structure was available on an atomic level of detail. Their pioneering work was rewarded the Nobel Prize in 1988.

The structural information then available has greatly contributed to the understanding of processes in photosynthesis, especially proton and electron transfer and pumping.

[1] Other bacterial photosynthesis processes exist, which use another protein, bacteriorhodopsin. Bacteriorhodopsin belongs to the same class of proteins as Rhodopsin, a protein involved in the visual process, another light-triggered process.

[2] James Allen wrote an interesting personal account of his time as a post-doc in the lab [7].

1.2 Proton transfer

Proton transfer (PT) is essential for the working of metabolic processes in cellular beings from bacteria to humans.

In many enzymes, local PT between the substrate and amino acids forms the basis for the general acid-base catalysis [8]. Long-range proton exchange between distant active sites has been recently proposed to be a mechanism for catalytic co-operativity [9].

In addition, PT plays an important role in bioenergetics. As illustrated above for the reaction centers, protons are *pumped* across the lipid membrane through proteins (such as the reaction centers described above) embedded in it [10–13] to create a proton concentration gradient. Another example for such proton pumps is cytochrome c oxidase in the respiration chain. These pumps can be driven by light or by electron transfer processes.

The *pump-turbine* combination can only work efficiently, if leaking of protons is kept to a minimum. The membrane itself acts an an *insulator*, and direct leakage of protons through the cell membrane is a very rare event The metabolism of cells requires a number of metabolites to transfer membranes. Specific channels exist for transporting both charged and non-charged particles but generally do not permit proton passage. For instance, aquaporins can pass water and other small, uncharged molecules [14–16], but do not conduct protons [17–19]. On the other hand, for example, the mechanism of gramicidin A, a protein antibiotic, is to introduce a channel into the cell membrane, which conduct protons, hence undermining this crucial process [20–22]. Another example where for bypassing the ATP synthase machinery is the thermogenin channel found in fat cells of newborns and in hibernating animals. Thermogenin is a switchable proton channel, which short circuits the proton circuit, generating heat.

In addition to the biological relevance of proton transfer, a number of technical applications, such as fuel cells [23], require well-controlled long-range PTs.

1.3 Investigating proton transfer

Major efforts from both the experimental [24, 25] and theoretical community [26–30] in the past several decades, have uncovered principles of basic mechanisms for localized PT in enzyme active sites. E.g., it is accepted that the rate of the proton transfer is largely modulated by electrostatic interactions [31]. Also, proton donor-acceptor dynamics, which is likely coupled to the overall fluctuation of the enzyme, plays an important role [32].

Long-range PTs, by contrast, are more challenging to understand at a quantitative level. It is generally accepted that long-range PT occurs through the help of water molecules, forming hydrogen-bonded "wires" [33, 34], in many cases with participation of titratable amino acid sidechains [35]. However, the precise transfer pathways and rate-limiting factors are often difficult to unravel due to the large number of solvent and protein residues involved.

Experimentally, mutagenesis experiments and kinetic measurements for the mutants can in principle be used to probe the transfer pathway and important interactions that regulate the transfer kinetics. Yet, mutation may introduce non-trivial perturbations. For instance, eliminating one amino acid as a proton carrier by mutation might not only affect the pathway connecting through this amino acid, but can also alter the proton affinities of surrounding residues, altering the overall energetics and kinetics. Hence, some ambiguity in the interpretation of the results obtained remain.

In the BRC, for example, even after a large number of experimental studies over more than three decades, the precise mechanism and sequence of events remain controversial (for reviews see, e.g., [11, 36–38]). Although the amino-acids participating in proton transfer are well established, their precise role and the pathways connecting them are not clear.

Theoretical analysis, therefore, is a complementary technique and can, eventually, provide mechanistic insights on an atomistic level. Indeed, a large number of theoretical studies have been applied to several systems involving long-distance PT, such as carbonic anhydrase [39–44], gramicidin [20, 21, 21, 45, 46], bacteriorhodopsin [47, 48], aquaporin [17–19], the synthetic LS2-channel [49] and cytochrome c oxidase [50–53]. These theoretical studies certainly provided valuable mechanistic information, but it is perhaps fair to state that a quantitative understanding has not been obtained in many of those systems. For instance, in gramicidin A [20–22] and aquaporin [17–19], studies from a number of groups using different simulation protocols proposed rather different rate-limiting events.

These controversies reflect the fact that a number of technical challenges have to be overcome for generally robust theoretical analysis of long-range PT. In this work, two of those issues will be addressed:

First, since a large number of bond-breaking and formation events are involved, an accurate description of the reaction energetics requires a quantum mechanical treatment. Special techniques have to be used to limit the computational cost, partitioning the system into multiple regions with different levels of description. Here, quantum mechanical/molecular mechanical coupling (QM/MM) was used. The treatment of the QM/MM frontier was and is the subject of dispute in the community. Inaccurate treatment can jeopardize the accuracy of the results.

Second, since a large number of shuttling groups are explicitly involved in the long-range PT, sufficient conformational sampling of the protein and solvent atoms at a relevant temperature is required for quantitative estimate of the transfer kinetics. In many previous studies minimum energy paths (MEP) were computed using the nudged elastic band (NEB) [54] or conjugated path refinement (CPR) [55] algorithms. These MEPs are useful for qualitative analysis but inadequate for a more quantitative comparison with experiments, unless coupled with methods for extensive sampling, see for instance [56]. With more demanding computational cost, potential of mean force (PMF) was evaluated for the PT in several studies [17–21, 45, 46, 49]. In this way the effect of thermal fluctuations can be included. *However*, one rather general problem concerns the choice of reaction coordinate in the PMF calculations. Most previous approaches either assume a specific pattern of transfer pathways (e.g., step-wise) or work best only for transfers across linear wa-

ter chains.

The long-term goal is to understand the mechanism of proton transfer involved in biochemical processes.

1.4 Overview of this work

The next two chapters introduce the methodological groundwork used for the work presented here:

In Chap. 2 the potentials used for describing biomolecules in this work are introduced. Methods used in this work to explore the potential of a molecular system are described in Chap. 3.

The following two chapters present technical developments to the goal of understanding long range proton transfer:

Chap. 4 gives a critical analysis of different QM/MM frontier treatments. A simple scheme is suggested for the description of charges at the QM/MM frontier and is compared to other approaches suggested in the recent literature.

A reaction coordinate for describing long-range proton transfer is introduced in Chap. 5. Its collective nature allows the calculation of potential of mean force without assuming *a priori* the pathway of the transfer, and they are sufficiently flexible to cope with non-linear PT pathways. Shortfalls of previous suggestions for reaction coordinates in linear proton wires and improvements are discussed.

As an application for the methods developed here, the study of a long-range proton transfer process in the BRC is presented in Chap. 6.

The conclusion in Chap. 7 summarizes the results and gives an outlook on future research this thesis has laid the groundwork for.

Chapter 2

Potentials

2.1 Overview

This part of the work explains how the energetics of molecular systems are described within this research. To treat the dynamics of a molecular system, deformations of the geometry need to be described. For this purpose, we need a suitable functional form which maps coordinate fluctuations onto an energy function. A multitude of different approaches exist today to achieve this. In this work atomistic descriptions are used to describe the structure of molecules, i.e., the atoms are treated explicitly. This is in contrast to other structural approaches which use coarser units of flexible subunits to describe flexibility of molecules [57].

The atomistic models used here can further be divided into those based on electronic structure (introduced in Sec. 2.1) and those based on a mechanical model of molecules (introduced in Sec. 2.2).

This work relies on the CHARMM27 force field [58] and the approximate density functional treatment SCC-DFTB [59] as well as their combination [60] as implemented in the CHARMM (Chemistry at HARvard Molecular Mechanics) program [61]. This combination has been applied to a variety of biological systems [62–69] and has shown to be particularly applicable for enzyme studies due to its speed and satisfactory accuracy for many types of reactions, especially proton transfer.

Due to the complex, inhomogeneous environment in biomolecular systems special care must be taken to correctly describe long range electrostatics for accounting for effects such as electrostatic shielding. The solvation and membrane environment plays an important role for the structure of proteins [70]. Different electrostatic treatments and the continuum electrostatics approach used here are described in Sec. 2.4.

Each of the methods described here, electronic structure methods, molecular mechanics, and continuum electrostatics have their strengths and weaknesses. To unite the strengths and overcome some of the weaknesses, the different models are combined in a multiple-length-scale approach. The system is divided up into multiple subsystems which are described at different levels of theory, as shown in Fig. 2.1. An electronic structure method is used for a small part of the system requiring the highest level of theory. The protein and a limited number of water molecules are treated explicitly using a molecular mechanics method. The electrostatic field of the

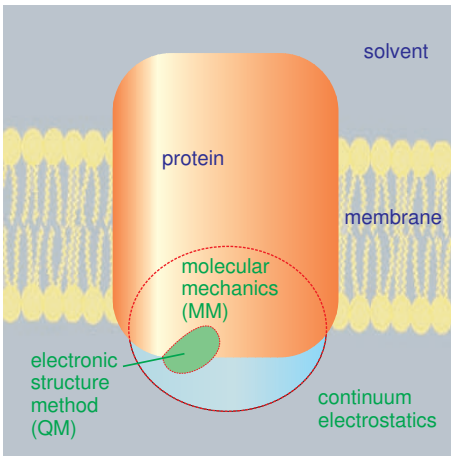


Figure 2.1: An example for multiscale partitioning as used in this work.

remaining protein and the effect of different electrostatic environments is treated using continuum electrostatics. The coupling of the different approaches is described in Sec. 2.3 and 2.4.

sectionQuantum mechanical treatment of the electronic structure

2.1.1 The quantum mechanical many-body problem

In the stationary case a system consisting of M nuclei and N electrons can be described using a Schrödinger-Equation of the type shown in Eq. 2.1.

$$\widehat{H}_{tot}(\{R\},\{r\})\Psi_{tot}(\{R\},\{r\}) = E^i\,\Psi_{tot}(\{R\},\{r\}) \tag{2.1}$$

Here $\widehat{H}_{tot}$ and Ψ_{tot} denote the Hamilton-Operator and many body wave function for both the nuclei and the electrons as a function of their respective coordinates, $\{R\}$ and $\{r\}$. The Eigenvalue to $\widehat{H}_{tot}$ is the total energy of the system and is denoted by E^i.

Born Oppenheimer (BO) separation

In the *Born Oppenheimer (BO) separation* it is assumed that electronic structure adiabatically adapts to changes in nuclear geometry. Hence the slow movement of the nuclei is decoupled from the fast movement of the electrons. For this purpose, the wavefunction of the entire system is decomposed into a product of nuclear $\Psi_N(\{R\})$ and electronic $\Psi_{el}(\{R\},\{r\})$ wavefunctions. The dependence of the electronic wavefunction on the nuclear coordinates is purely parametrical and is hence written in the following as:

$$\Psi(\{r\}) = \Psi_e(\{R\},\{r\}) \tag{2.2}$$

Similarly, the Hamilton operator for a molecular system can be decomposed into a sum:

$$\widehat{H}_{tot}(\{R\},\{r\}) = \underbrace{\widehat{T}_N + \widehat{V}_{NN}}_{\widehat{H}_N\{R\}} + \underbrace{\widehat{T}_e + \widehat{V}_{Ne} + \widehat{V}_{ee}}_{\widehat{H}_e(\{R\},\{r\})} \tag{2.3}$$

Here, $\widehat{T}$ and $\widehat{V}$ are the kinetic and potential energy operators for nuclei (N) and electrons (e) respectively. The interaction energy between nuclei-nuclei, nuclei-electrons and electrons-electrons are described by V_{NN}, V_{Ne} and V_{ee}. We hence obtain two Schrödinger equations for electrons and nuclei:

$$\widehat{H}_e\Psi_e(\{r\}) = \left(\widehat{T}_e + \widehat{V}_{Ne} + \widehat{V}_{ee}\right)\Psi_e(\{r\}) = E_e\,\Psi_e(\{r\}) \tag{2.4}$$

$$\widehat{H}_N\Psi_N(R) = \left(\widehat{T}_N + \widehat{V}_{NN} + E_e\right)\Psi_N(\{R\}) = E_N\,\Psi_N(\{R\}) \tag{2.5}$$

In this work nuclei are treated classically. Thus the quantum mechanical many body problem reduces to solving the electronic problem (Eq. 2.4). The subscript "$_e$" will be dropped in the following.

$$\widehat{H}\Psi(\{r\}) = E\,\Psi(\{r\}) \tag{2.6}$$

Though the complexity starting Eq. 2.1 has been reduced significantly by the use of the BO-separation, we should not forget that we are still dealing with a many-body problem of interacting electrons. Two major classes of approaches to this problem exist. The Hartree-Fock (HF) method and post-Hartree Fock methods [71] deriving from it and density functional theory (DFT). The later is the basis for the DFTB method (Sec 2.1.3) used in this work. Following is a brief outline of theory.

In the following work quantum mechanical treatment will always refer to quantum mechanical treatment of the electronic structure unless otherwise noted. Although this terminology is not precise, it is prevalent in the literature, especially for describing multiscale approaches.

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.

2.1.2 Density functional theory

The Hohenberg-Kohn theorems

density functional theory (DFT)

The basis of the *density functional theory (DFT)* are the theorems of Hohenberg and Kohn, who showed, that the ground state energy E_0 of an electronic system is only determined by its electron density n_0 [72]. Using the density functional theory, the complexity of the problem of an N-electron problem is reduced dramatically by decreasing the number of coordinates from 3N to 3.

$$E_0 = E[n_0] \tag{2.7}$$

universal density functional

The exact functional mapping the electron density to the energy, however, is unknown. The *universal density functional* $F[n]$ maps the electron density n to its energy $F[n]$, not taking into account the external potential V_{ext}.

$$F[n] = T[n] + V_{ee}[n] \tag{2.8}$$

where $T[n]$ and $V_{ee}[n]$ are the electronic kinetic energy and electron-electron interaction functionals. The theorems of Hohenberg-Kohn tell us that the variational principle holds for an approximated density for the total electronic energy:

$$E[n] = F[n] + \int V_{ext}(\boldsymbol{r})\,n(\boldsymbol{r})d\boldsymbol{r} \geqslant E_0 \tag{2.9}$$

where for the exact ground state density $n = n_0$ the equal sign holds.

The first approximations to density functionals assumed a homogeneous, non-interacting electron gas. The energy functional of Thomas, Fermi and Dirac used pure density functionals for the kinetic T, coulomb J and non-classical exchange energy K contributions:

$$F[n] = T[n] + J[n] + K[n] \tag{2.10}$$

The greatest problem with this approach to density functional theory was the insufficient accuracy in describing the kinetic energy [73]. Part of these problems were overcome by Kohn and Sham [74].

The Kohn-Sham equations

Kohn and Sham assumed a non interacting electron gas of N electrons. The wave function of such a system can be written in form of a single Slater determinant of one-particle wave functions and associated occupation numbers n_i. The total electron density can be expressed as:

$$n(\boldsymbol{r}) = \sum_{i=1}^{occ} n_i |\psi_i(\boldsymbol{r})|^2 \qquad \text{with} \qquad N = \sum_{i=1}^{occ} n_i \tag{2.11}$$

In this case, the kinetic energy and total energy of the system is given by:

$$F[n] = t[n] = \sum_{i=1}^{occ} n_i \left\langle \psi_i(\boldsymbol{r}) \left| \widehat{t} \right| \psi_i(\boldsymbol{r}) \right\rangle = \sum_{i=1}^{occ} n_i \left\langle \psi_i(\boldsymbol{r}) \left| -\frac{\Delta}{2} \right| \psi_i(\boldsymbol{r}) \right\rangle \tag{2.12}$$

For an interacting system, this expression is no longer correct. The kinetic energy expression can no longer be written in this form and the wave-function can generally no longer be written in form of a single Slater determinant. Kohn and Sham hence suggested a reference system of non-interacting electrons which reproduces the electron density of the fully-interacting system. The total energy of the system then reads as follows:

$$E^{KS} = \sum_{i}^{N} n_i \left\langle \psi_i \left| -\frac{\Delta}{2} \right| \psi_i \right\rangle + \int v_{ext}(\boldsymbol{r}) n(\boldsymbol{r}) d\boldsymbol{r} + J[n] + E_{xc}[n] \tag{2.13}$$

The energy contribution of the coulomb functional $J[n]$ is computed as:

$$J[n] = \frac{1}{2} \iint \frac{n(\boldsymbol{r}) n(\boldsymbol{r'})}{|\boldsymbol{r} - \boldsymbol{r}_0|} d^3\boldsymbol{r} d^3\boldsymbol{r'} \tag{2.14}$$

exchange-correlation energy

The new energy term introduced in Eq. 2.13, the *exchange-correlation energy* E_{xc} includes the non-classical exchange energy term, as well as the correlation energy and a correction term for the kinetic energy, accounting for the actual dynamical correlated movement of interacting electrons. The exchange-correlation energy is defined as the difference between the correct Hohenberg-Kohn energy functional $F[n]$ and the Kohn-Sham energy terms:

$$E_{xc}[n] = F^{HK} - t[n] - J[n] \tag{2.15}$$

Various suggestions were made for E_{xc} [75]. However, the *exact* functional form is unknown.

Using the variational principle a one-electron Hamiltonian can be derived for solving the one-electron Kohn-Sham eigenvalue problem, with eigenvalues ϵ_i:

$$\widehat{h} = -\frac{\Delta}{2} + v_{ext} + \int \frac{n(\boldsymbol{r'})}{|\boldsymbol{r} - \boldsymbol{r}_0|} d^3\boldsymbol{r'} + v_{xc}[n] \tag{2.16}$$

$$\widehat{h}\,|\psi_i\rangle = \epsilon_i\,|\psi_i\rangle \tag{2.17}$$

The exchange-correlation potential $v_{xc}[n]$ introduced in the single electron Hamiltonian (2.16) is the functional derivative of the exchange-correlation energy $E_{xc}[n]$:

$$v_{xc}[n] = \frac{\delta E_{xc}[n]}{\delta n} \tag{2.18}$$

Using the sum of eigenvalues ϵ_i an alternate expression for the total energy (2.13) can be obtained, which will be needed for the derivation of the DFTB energy expression (vide infra):

$$E_{tot}^{KS} = \sum_i^N n_i \left\langle \psi_i \left| -\frac{\Delta}{2} + v_{ext} + \frac{1}{2}\int \frac{n(\boldsymbol{r'})}{|\boldsymbol{r} - \boldsymbol{r}_0|} d^3\boldsymbol{r'} \right| \psi_i \right\rangle + E_{xc}[n] + V_{nn} \tag{2.19}$$

Here, V_{nn}, the nuclear-nuclear interaction potential was included in the energy term.

With density functional theory the complexity of the electronic problem could be reduced in comparison to post-Hartree-Fock methods. Still, using modern density functionals such as the B3LYP hybrid functional [76], the accuracy of results is respectable in comparison [77].

Despite these advantages, computational cost is still prohibitively high for very large systems or for extended molecular dynamics simulations. Even with QM/MM coupling techniques (Sec. 2.3) and parallelization, only short trajectories can be obtained for statistical sampling, see for instance [78].

2.1.3 Density functional theory based tight binding

One approach to reduce the cost of density functional calculations was the introduction of approximate density functional theory based frameworks, such as tight binding approaches. Approximations include (partially) dropping the self-consistency requirement of DFT, dropping higher-center integrals, tabulating the required integrals and limiting the size of the basis set to a minimal basis. In the development of the *density-functional theory based tight binding (DFTB)* [79] significant speed improvements have been made while retaining the accuracy of DFT in many cases. Overall DFTB and SCC-DFTB (vide infra) yield an excellent accuracy/ computational cost ratio.

density-functional theory based tight binding (DFTB)

An overview of developments from a historical perspective can be found in [80], while accounts of recent developments and applications can be found elsewhere [62, 81].

In contrast to earlier tight binding methods, the matrix elements of the Hamiltonian and the other terms, such as the repulsive term (vide infra) are not parametrized to experimental data, but instead are obtained from fully self-consistent density functional computations.

The idea behind DFTB is the expansion of the electron density around a reference density n_0 as suggested by Eschrig and Seifert [82] and others [83]:

$$n(\boldsymbol{r}) = n_0(\boldsymbol{r}) + \delta n(\boldsymbol{r}) \tag{2.20}$$

Then, the total energy in Eq. 2.19 can be expanded, obtaining the following expression, neglecting for the moment the expansion of the exchange-correlation functional $E_xc[n]$:

$$\begin{aligned} E_{tot}[n_0 + \delta n] = &\sum_i^N n_i \left\langle \psi_i \left| -\frac{\Delta}{2} + v_{ext} + \int \frac{n_0(\boldsymbol{r'})}{|\boldsymbol{r} - \boldsymbol{r'}|} d^3\boldsymbol{r'} + v_{xc}[n_0] \right| \psi_i \right\rangle \\ &- \frac{1}{2} \iint \frac{n_o(\boldsymbol{r'})(n_o(\boldsymbol{r}) + \delta n(\boldsymbol{r}))}{|\boldsymbol{r} - \boldsymbol{r'}|} d\boldsymbol{r} d\boldsymbol{r'} \\ &- \int v_{xc}[n](\boldsymbol{r}) \left(n(\boldsymbol{r}) + \delta n(\boldsymbol{r})\right) d\boldsymbol{r} \\ &+ \frac{1}{2} \iint \frac{\delta n(\boldsymbol{r'})(n_o(\boldsymbol{r}) + \delta n(\boldsymbol{r}))}{|\boldsymbol{r} - \boldsymbol{r'}|} d\boldsymbol{r} d\boldsymbol{r'} + E_{xc}[n] + V_{nn} \end{aligned} \tag{2.21}$$

Introducing the second order Taylor expansion of the exchange-correlation functional, we obtain the following expression, where all the terms linear in δn have canceled out:

$$\begin{aligned} E_{tot} = &\sum_i^N n_i \left\langle \psi_i \left| H^0 \right| \psi_i \right\rangle \\ &- \frac{1}{2} \iint \frac{n_o(\boldsymbol{r'}) n_o(\boldsymbol{r})}{|\boldsymbol{r} - \boldsymbol{r'}|} d\boldsymbol{r} d\boldsymbol{r'} - \int v_{xc}[n_0](\boldsymbol{r}) n(\boldsymbol{r}) d\boldsymbol{r} + E_{xc}[n_0] + V_{nn} \\ &+ \frac{1}{2} \iint \left[\frac{1}{|\boldsymbol{r} - \boldsymbol{r'}|} + \left(\frac{\delta^2 E_{xc}[n]}{\delta n(\boldsymbol{r}) \delta n(\boldsymbol{r'})} \right)_{n=n_0} \right] \delta n(\boldsymbol{r}) \delta n(\boldsymbol{r'}) d\boldsymbol{r} d\boldsymbol{r'} \end{aligned} \tag{2.22}$$

The total energy terms can be regrouped as:

$$E_{tot} = E_{BS} + E_{rep} + E_{2nd} \tag{2.23}$$

Omitting the second order corrections E_{2nd}, i.e., the last term in Eq. 2.23 yields the Hamiltonian for the DFTB method, which gives reasonable results for systems where charge transfer can be neglected. More details on the second order corrections will be given below. The *band structure energy* E_{BS} is defined as the sum over all eigenvalues:

band structure energy

$$E_{BS} = \sum_i^N n_i \epsilon_i \tag{2.24}$$

repulsive energy

The *repulsive energy* E_{rep} comprises a collection of terms which are fully repulsive:

$$E_{rep} = -\frac{1}{2}\iint \frac{n_o(\boldsymbol{r'})n_o(\boldsymbol{r})}{|\boldsymbol{r}-\boldsymbol{r'}|}d\boldsymbol{r}d\boldsymbol{r'} - \int v_{xc}[n_0](\boldsymbol{r})n(\boldsymbol{r})d\boldsymbol{r} + E_{xc}[n_0] + V_{nn} \tag{2.25}$$

Further, the repulsive energy depends only on the reference density n_0. If n_0 is the density for the neutral atoms, then the nuclear-nuclear repulsion is shielded. Under the assumption of only local contributions to the exchange correlation potential, E_{rep} can then be assumed to be short-ranged and purely pairwise, depending only on atom-atom distances. The repulsive energy term is determined using the difference between the binding curve of a reference system determined in a fully self consistent calculation and the band-structure term obtained using DFTB, and possibly E_{2nd} for SCC-DFTB parametrizations.

$$\begin{aligned} E_{rep}(R_{AB}) = E^{KS}_{tot}(R_{AB}) - \sum_{i}^{occ}\left[n_i\epsilon_i\right](R_{AB}) \\ - E_{2nd}(R_{AB}) \qquad \text{for SCC-DFTB} \end{aligned} \tag{2.26}$$

where R_{AB} is the interatomic distance.

Self-consistent-charge DFTB

For the description of molecules with significant charge transfer between the atoms, the second order terms have to be taken into account to correctly describe charge transfer and molecular properties such reaction energetics, geometries or molecular vibrations [59]. These terms are included in the *self consistent charge (SCC)* extension of the DFTB formalism.

self consistent charge (SCC)

The second order correction E_{2nd} is dependant on the charge fluctuations. The density fluctuations can be developed into atom centered contributions Δq_A. The following functional form was suggested for E_{2nd} [59]:

$$E_{2nd} = \frac{1}{2}\sum_A\sum_B \gamma_{AB}(R_{AB})\Delta q_A \Delta q_B \tag{2.27}$$

The function $\gamma_{AB}(R_{AB})$ is the first term of multipole expansions in the density fluctuations. For large interatomic separations $R_{AB} \rightarrow \infty$ $\gamma_{AB}(R_{AB}$ it describes the coulomb interaction between partial charges on atoms A and B: $\gamma_{AB}(R_{AB}) \rightarrow 1/R_{AB}$. The value of γ_{AA} is given by the Hubbard parameters of atoms A: U_A. The Hubbard parameter is directly related to the chemical hardness η_A with $U_A = 2\eta_A$. For a detailed derivation of γ see [84].

Basis set

In DFTB and SCC-DFTB the electron density and Kohn-Sham wave functions are expanded into an atom-centered, contracted basis functions consisting of Slater-type functions:

$$\psi_i(\boldsymbol{r}) = \sum_\nu c_{\nu i}\phi^A_\nu(\boldsymbol{r}-\boldsymbol{R}_A) \qquad \nu \in A \tag{2.28}$$

The basis is minimal and only valence electrons are included in the description. From the expansion coefficients $c_{\nu i}$ the Mulliken atomic charges needed for the self-consistent charge treatment can be obtained, using elements of the overlap matrix **S**:

$$q_A = \sum_i^{occ} n_i \sum_{\mu \in A} \sum_\nu c_{\mu i} c_{\nu i} S_{\mu\nu} \tag{2.29}$$

Hamiltonmatrix elements

Using the variational principle, the coefficients of the expansion in Eq. 2.28 can be obtained as solutions of the secular equation:

$$\sum_\nu c_{\nu i} \left(H_{\mu\nu} - \epsilon_\nu S\mu\nu \right) = 0 \qquad \forall i \tag{2.30}$$

The overlap matrix **S** can be directly computed from the given basis set. For the matrix elements of the Hamilton matrix **H** of zeroth order further simplifications are introduced:

$$H^0_{\mu\nu} = \left\langle \phi^A_\mu \left| H^0 \right| \phi^B_\nu \right\rangle = \begin{cases} \epsilon^{\text{free atom}}_\mu & \text{if } \mu = \nu \\ \left\langle \phi^A_\mu \left| \widehat{T} + \widehat{V}^A_0 + \widehat{V}^B_0 \right| \phi^B_\nu \right\rangle & \text{if } A \neq B \\ 0 & \text{otherwise.} \end{cases} \tag{2.31}$$

This formulation of the Hamilton matrix **H** implies the neglect of crystal field terms ($< \phi^A_\mu | \widehat{V}^B_0 | < \phi^A_\mu >$) and three-center integrals ($< \phi^A_\mu | \widehat{V}^C_0 | < \phi^B_\nu >$). The matrix elements of **S** and **H** are tabulated for a range of distances of nuclear separations during the parametrization. During the actual computation table lookups and interpolations are used.

2.2 Molecular force fields

2.2.1 Concept

molecular force fields
molecular mechanics methods

An alternative model to electronic structure based methods are *molecular force fields* - also termed *molecular mechanics methods*. While still operating on an atomistic level these methods neglect the electronic structure of matter. The underlying assumption is BO-separability, i.e., that the the energy can be written as a function of the nuclear coordinates only. The idea is to describe a molecular system as atom-centered mass-points connected by an elastic framework and to find an empirical functional which describes the energy penalties associated with deformations of the framework.

atom types

Atoms are classified into *atom types*, where not only the respective element is important for its classification, but also its binding situation. These types dictate, how atoms in the framework interact. For instance, the potential for stretching an O-C bond in an alcohol is different from stretching an O=C bond in a ketone, which is different from stretching an O=C bond in a peptide.

transferability

On one hand, one general aim of force fields parametrization is great *transferability*, meaning that the parameters should be applicable to a different molecule showing a similar bonding situation. For instance, force fields such as the MM force fields of Allinger et al. [85, 86] are fairly general with atom types in most cases only determined by the hybridization of the atom. Generally, even with more specialized force fields, often parameters can be adapted directly to other molecules (parametrization by analogy).

In other cases sacrificing transferability by reparametrization for specific cases can give a great improvement in *precision* [87].While general force fields are widely applicable, specialized force fields for instance for proteins and other biomolecules [58, 88] can offer higher precision.

precision

In terms of transferability biomolecules should represent a limited amount of effort needed for parametrization. Thinking of biopolymers, such as proteins, there is only a limited number of monomers commonly used. Nature only knows 21 proteinogen amino acids common to all species. Despite this seemingly small diversity, there are still improvements being made to the parametrizations.

Another way of improving the description by force fields is to introduce new functional forms. Molecular force fields are empirical that is there is no *correct* functional form. Various functional forms and parametrizations have been introduced as reviewed in [89] and new force field terms are still being developed to tackle known problems, e.g., [90].

2.2.2 The CHARMM forcefield

In the work presented here the CHARMM27 [58] force field is used. The functional form is similar to that used in other force fields such as the AMBER force field [88], or the MM force fields of Allinger et al. [85, 86].

The energy expression for such a force field is written as the sum of of *inter- and intramolecular energy contributions* (Eq. 2.32). Here, the term intramolecular refers to terms derived from the topology of the molecule as defined by the *Lewis*-structure. In contrast intermolecular contributions are those not mediated through the network of bonds but instead through the general spacial arrangement. Intermolecular contributions are electrostatic and van der Waals interactions.

inter- and intramolecular energy contributions

$$E = E_{intra} + E_{inter} \tag{2.32}$$

Bonded interactions

A natural choice of coordinates for the intramolecular interactions is one which is directly derived from the topology of the molecule using bond lengths, angles, dihedral angles and improper angles as depicted in Fig. 2.2a)-d). The intramolecular potential describes the variation of energy connected to deviations from given geometrical parameters which are termed *natural parameters* [1]:

natural parameters

[1] For a discussion of the term natural parameters vs. equilibrium parameters see the review by Jensen [91]

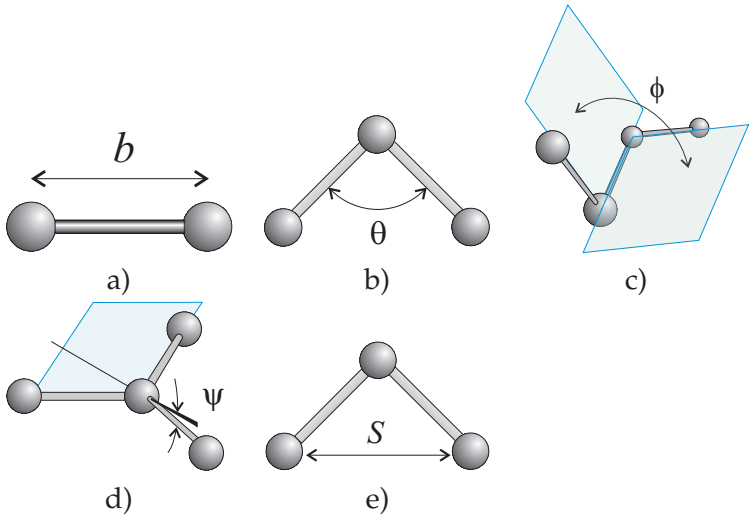


Figure 2.2: Relevant intramolecular geometry measures used in molecular force fields: a) bond lengths b) bond angles c) dihedral angles d) improper angles e) Urey-Bradley.

harmonic approximation

Small deformations from the the natural bondlength b_0 of a given bond are described using a *harmonic approximation* to the bond potential. The associated force constant is denoted by k_b:

$$E_{bond} = \frac{1}{2} k_b (b - b_0)^2 \tag{2.33}$$

More rigorous analytical forms exist for describing the energetics of bond separation such as the Morse potential which also in principle possess the correct asymptotic behavior for bond dissociation. Another approach to describe anharmonicities of the bond stretching is employed in the MM3 and MM4 force fields of Allinger et al. [85, 86], which also include cubic and quartic terms in their series expansion.

However, the harmonic approximation gives a reasonable qualitative and quantitative description for *small* deviations from the equilibrium bond length. For dealing with larger fluctuations, a different functional form has to be chosen or an electronic structure method has to be used, possibly in a multiscale approach (Sect. 2.3).

bond angle deformations

Similar approximations can be made for the description of *bond angle deformations* (Fig. 2.2b), where the deviation of the bond angle θ from its natural value θ_0 is described using an harmonic force term with force constant k_θ.

torsional deformations

The description of the *torsional deformations* using the dihedral coordinates ϕ (Fig. 2.2c) differs as the function is required to satisfy periodicity. Here, cosine terms using different periodicities n, amplified by the respective force constants $k_{n\phi}$ are employed to describe the energetics of deviation from the natural parameters δ_n. Different periodicities of the potentials have to be considered based on topology to describe more complex potentials.

out-of plane deformation terms

For some binding situations deviations from planarity have to be described in *out-of plane deformation terms* e.g., for the out-of-plane motion of the oxygen in a carbonyl compound. The angle ψ that the bond of one atom to the central atom deviates from

the plane of the remaining three atoms is used to define an harmonic term using the force constant k_ψ and natural value ψ_0.

The CHARMM force field further includes a force field term which is needed for correct vibrational behavior in some cases e.g., for H-C-H angles. In this *Urey-Bradley term* the distance S between the outer atoms connected by an angle term are used to define another harmonic potential with force constant k_{UB} (Fig. 2.2e).

Urey-Bradley term

The resulting energy term for the intramolecular contribution to the total energy reads as [2] :

$$\begin{aligned} E_{intra} = & \sum_{bonds} \frac{1}{2} k_b (b - b_0)^2 + \sum_{angles} \frac{1}{2} k_\theta (\theta - \theta_0)^2 \\ + & \sum_{torsions} \sum_{n} k_{n\phi} (1 + cos(n\phi - \delta_n)) + \sum_{impropers} \frac{1}{2} k_\psi (\psi - \psi_0)^2 \\ + & \sum_{\text{Urey-Bradley}} \frac{1}{2} k_{UB} (S - S_0)^2 \end{aligned} \tag{2.34}$$

Nonbonded interactions

As described above, nonbonded interactions are not mediated through the network of bonds but instead through space. Nonbonded interactions consist of contributions from electrostatic and van der Waals interaction. To facilitate parametrization, nonbonded interactions are excluded for atoms connected by bond or angle and in some cases dihedral terms.

For modelling electrostatic interactions in the simplest case the atoms i,j are regarded as point charges q_i,q_j separated by a distance r_{ij}. The interaction is the described according to Coulomb's law, where ϵ is the electric field constant of the medium:

$$E_{elec} = \frac{1}{4\pi\epsilon} \sum_{\substack{nonbonded, \\ I<J}} \frac{q_I q_J}{r_{IJ}} \tag{2.35}$$

A more detailed description and more sophisticated models for electrostatic treatment are given in Sect. 2.4.

The philosophy behind the CHARMM parameter sets is *interaction based parametrization*. This means that the point charges q_i for the atoms are parametrized to reproduce both interactions energies as well as structures, either using experimental or *ab-initio* data. The probe used in the parametrization of the CHARMM27 force field is the TIP3P water model [92].

interaction based parametrization

In contrast, other force fields use charges derived from ab-initio calculations such as condensed charges using Mulliken populations or electrostatic field based charge models [93–95].

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

The functional form for van der Waals interaction is:

$$E_{vdW} = \sum_{\substack{nonbonded,\\ I<J}} 4\epsilon_{IJ} \left[\left(\frac{\sigma_{IJ}}{r_{IJ}} \right)^{12} - \left(\frac{\sigma_{IJ}}{r_{IJ}} \right)^{6} \right] \tag{2.36}$$

The parameters entering the van der Waals contribution to the energy are the diameter of collision σ and the well depth ϵ. For heteroatomic interactions the parameters are computed using the combination rules $\sigma_{ij} = 1/2(\sigma_i + \sigma_j)$ and $\epsilon_{ij} = \sqrt{\epsilon_i \epsilon_j}$

London dispersive interactions

The term contribution to r^{-6} in 2.36 is attractive and stems from *London dispersive interactions*. London showed how these interactions between molecules can be understood using Drude oscillators [91]. London dispersion interactions are due to instantaneous fluctuations in the electron density of neighboring electron clouds giving rise to dipole - induced dipole interactions. Here, the higher terms in the expansion of interactions of higher multipoles are neglected.

exchange interaction

The repulsive, shorter ranged contribution in 2.36 can be understood in terms of - *exchange interaction* and nuclear repulsion. The choice of functional form of r^{-12} is purely empirical[3].

2.3 Quantum mechanical – molecular mechanical coupling

2.3.1 Underlying idea

Quantum mechanical (QM) approaches can describe the breaking and the formation of chemical bonds and hence are required for studying chemical reactivity. Many reactions of interest involve crucial contributions from the environment, which can be solvent or a macromolecule. Various approaches for including the effect of the environment on chemical reactivities have been suggested and implemented. For example, linear scaling techniques [96] are available in different theoretical frameworks and have been successfully applied to realistic biomolecular systems [97]. The potential of these full QM methods is enormous considering the speed gain of computational resources in recent years [98].

Still, many problems in Molecular Modelling in particular in biophysics relevant areas are too large to be treated using electronic structure methods. Even with semi-empirical treatments [99, 100] or approximate density functional theory [59] scaling is still unfavorable. The effort for solving an N electron problem still increases steeply and is in $O(N^3)$. Hence, performing a sufficient amount of statistical sampling for molecular dynamics in realistic systems remains difficult.

The computational performance for molecular force fields are more favorable. However, generally they are not suitable for describing bond breaking and formation;

[3] Other functional dependencies such as r^{-9} have been suggested. One reason why the r^{-12} form has prevailed is for numerical efficiency in earlier times. Computing odd powers of r requires taking the square root of the squared differences of the coordinate components whereas only including even powers does not require this step for neither energy nor forces.

some notable exceptions, so called reactive force fields, exist such as the PM6 water force field allowing the description of water hydrolysis and proton transfer [101]. Hence, quantum mechanical treatment is required for studying chemical reactivity.

To unite the strengths of both methods, an intuitive, yet powerful approach is to partition the system in a multi-scalar approach, thereby treating different subsystems with different levels of theory. One possibility here is the *hybrid quantum mechanical/molecular mechanical (QM/MM) approach* [102–104], in which the reactive part is described with QM and the remaining system by MM. The underlying assumptions are that the changes in the electronic structure are localized and limited to a select number of atoms, whose potential described using a quantum mechanical method, and that the response of the environment is described with sufficient accuracy using a force field. Since the initial introduction of the QM/MM idea by Warshel and Levitt nearly 30 years ago numerous studies have been conducted examining both performance of different schemes as well as interesting applications to various solution and enzyme systems [27, 31, 105, 106].

hybrid quantum mechanical/molecular mechanical (QM/MM) approach

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.

In commonly used QM/MM schemes the Hamilton 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, $\widehat{H}_{QM/MM}$,

$$\widehat{H} = \widehat{H}^{QM} + \widehat{H}^{MM} + \widehat{H}^{QM/MM} \tag{2.37}$$

A different functional form is used in so called subtractive schemes [107].

The precise expression of $\widehat{H}_{QM/MM}$ varies but generally has contributions from electrostatic, van der Waals, bonded interactions and possibly additional constraints [102],

$$\widehat{H}^{QM/MM} = \widehat{H}^{QM/MM}_{elec} + \widehat{H}^{QM/MM}_{vdW} + \widehat{H}^{QM/MM}_{bonded} + \widehat{H}^{QM/MM}_{cons} \tag{2.38}$$

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 [108].

2.3.2 Embedding schemes

QM/MM embedding schemes can be roughly classified into two large categories by the kind of interactions included in $\widehat{H}^{QM/MM}$ (beyond covalent terms at the frontier, $\widehat{H}^{QM/MM}_{bonded}$ and constraints $\widehat{H}^{QM/MM}_{cons}$):

1. *Mechanical embedding* This crudest way of treating the embedding only treats

Mechanical embedding

steric interactions between the QM and MM partition using van der Waals interactions.

Electrostatical embedding

2. *Electrostatical embedding* This approach goes beyond the mechanical embedding by also including electrostatic interactions between the QM and MM partition.

The first choice is well suited for describing situations in which the main effect of the environment is directing the reaction by steric interactions. Mechanical embedding approaches such as the ONIOM model [107] have been successfully used for describing systems in catalysis [107], biomolecular relevant molecules, e.g., [109] and the description of surface processes, e.g., [110].

The shortcoming of this approach is that is does not allow for polarization. According to the prevalent model of enzyme catalysis, polarization plays an essential role. A recent example for the importance of polarization in biomolecules is given in [111].

The electrostatical embedding schemes can further be divided into schemes which only allow for polarization of the QM zone and those which allow backpolarization of the MM zone. The latter requires a polarizable force field, which allows for a variation of charges on the atoms in the MM zone. Although some suggestions for polarizable force fields are available [112, 113], there are issues with transferability and stability. Hence in this work, polarization is limited to the QM atoms.

For the purpose of computing electrostatic interactions between the QM and MM atoms the charge distribution in the MM zone in most cases is assumed to be described by point charges on the individual atoms. The electrostatic part of $\widehat{H}_{QM/MM}$ hence has the following functional form:

$$\widehat{H}_{elec}^{QM/MM} = \sum_{A \in MM} q_A \int \frac{\rho^{QM}(\boldsymbol{r})}{|\boldsymbol{r} - \boldsymbol{r_A}|} d\boldsymbol{r} \tag{2.39}$$

2.3.3 QM/MM and SCC-DFTB

In the original implementation[114], $\widehat{H}_{QM/MM}^{elec}$ has the form of Coulomb interaction between the MM point charge Q_A and the Mulliken charge Δq_B on the QM atom [114].

$$\widehat{H}_{elec,pointcharges}^{QM/MM} = \sum_{A \in MM} \sum_{B \in QM} \frac{Q_A \, \Delta q_B}{|\boldsymbol{r}_A - \boldsymbol{r}_B|} \tag{2.40}$$

The corresponding contribution to the Hamiltonian matrix elements is:

$$\mathbf{H}_{\mu\nu}^{elec,pointcharges} = \frac{1}{2} \sum_{A \in MM} \left(\frac{Q_A}{|\boldsymbol{r}_C - \boldsymbol{r}_A|} + \frac{Q_A}{|\boldsymbol{r}_D - \boldsymbol{r}_A|} \right) \quad \mu \in C, \nu \in D \tag{2.41}$$

2.3.4 Problems in the QM/MM approach

An issue that has been repeatedly raised concerns the treatment of the QM/MM boundary, for recent discussions see [115, 116]. 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 artifactual effects on the electronic structure of the QM region.

Issues concerning the precise treatment of the QM/MM boundary for these cases have been repeatedly raised over the last years. An overview of the current discussion as well as the evaluation of different schemes in the SCC-DFTB/CHARMM framework will be given in Chap. 4.

2.4 Long-range electrostatics

2.4.1 Description of electrostatics in molecular dynamics simulations

The computation of non-bonded interactions in molecular dynamics simulations dominates the cost in large molecules which was one of the first reasons the simplification of these terms was considered.

Initial suggestions were to not consider interactions beyond a certain cutoff distance. While simple truncation schemes lead to severe artifacts due to the associated discontinuities, schemes including smoothening functions such as switching or shifting for the potential or forces are widely used [117]. While using cutoffs most certainly is appropriate for the short-ranged van der Waals interactions, the description of electrostatic interactions may suffer significantly, especially for processes with large scale reorganization or long-range charge shifts. Stote et. al pointed out that even the description of the structure of metal-containing enzymes may suffer if long-range-electrostatics is neglected [118].

extended electrostatics model

A suggestion which circumvents these caveats is the *extended electrostatics model* [119], where interactions beyond a certain cutoff are treated using a multipolar expansion including dipole and quadrupole terms.

electrostatic shielding

Apart from these technical and scaling issues, a more severe point is the description of the inhomogeneous environment (aqueous solvent, possibly membrane) in biomolecular systems and its effect on *electrostatic shielding*. Different environments with associated different dielectric constants strongly modulate the Coulombic interactions and fields created. The reaction field of the solvent with its high dielectric constant screens out the electric field generated by charge distributions. For instance, protein helix dipoles, which are argued to play an important role for effects such as excitation spectra, reaction barriers and pK_a values, are dampened out effectively for some solvent-protein geometries [120].

Several recent studies [68, 121–124] showed that appropriate treatment of electro-

statics is utterly important to molecular dynamics simulations. The issue is most serious for the simulation of processes that involve significant change in the charge distribution, such as oxidation-reduction reactions and long-range proton transfers. Inappropriate treatment of electrostatics may not only affect the quantitative aspect of the result, but also produce qualitative changes in the behavior of biological systems. Although electrostatic interactions are known to be crucial to enzyme catalysis [125], yet electrostatics are often treated in highly approximate manner in QM/MM simulations. Partial charges for charged residues on the surface of the enzyme are either quite arbitrarily set to zero or scaled down [126] according to Poisson-Boltzmann calculations based on a *single* conformation. Moreover, in stochastic boundary simulations [127], atoms outside of the buffer region are often deleted and the reservoir region is essentially vacuum; some studies keep the reservoir atoms but the dielectric screening effect due to the bulk solvent is ignored. All these rather *ad hoc* approximations may work for reaction path type of analysis for very localized chemical reactions, but they are unlikely to be robust for simulations with extensive conformational sampling.

For example, water structure in the active site of enzyme carbonic anhydrase was found to differ significantly with different electrostatic models used in QM/MM simulations [122]. In the study of aquaporins, we showed that the water structure in the lumen found both in experiment and large-scale molecular dynamics simulations can only be reproduced in a slab model, if proper long-range electrostatics treatment is used [124].

Ewald summation

The most robust electrostatic model up to date involves *Ewald summation* (or related numerical improvements) with the periodic boundary condition [128]. However, the necessity of including a large number of explicit solvent molecules to avoid artifacts in such simulations limits its use in QM/MM simulations of large systems [121, 129]. Since the systems of interest in the context of proton pumping are often rather large, it is desirable to adopt schemes that avoid periodic boundary condition. In this regard, the *generalized solvent boundary potential (GSBP)* approach proposed by Roux and co-workers [130] is very attractive. It treats a relatively small region of the protein-solvent-membrane system with microscopic details, with the contribution from the rest 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 extension of GSBP to a QM/MM framework and SCC-DFTB was recently described in the literature [122]. In the following, the GSBP approach in its original form [130] and the implementation in the QM/MM framework is briefly reviewed.

generalized solvent boundary potential (GSBP)

2.4.2 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 boundaries between different dielectric environments.

To quantitatively treat these effects we have to consider the electrostatic contribution

to the solvation free energy:

$$\Delta W_{elec}(\boldsymbol{r}) = \frac{1}{2}\sum_A q_A \phi_{rf}(\boldsymbol{r}_A) \tag{2.42}$$

Here $\phi_{rf}(\boldsymbol{r})$ is the *reaction field* at the position $\boldsymbol{r}$ of the charge q_A. The reaction field is defined as the difference between the electrostatical potential of the solvated system $\phi_s(\boldsymbol{r})$ and the system in vacuum $\phi_v(\boldsymbol{r})$: reaction field

$$\phi_{rf}(\boldsymbol{r}) = \phi_s(\boldsymbol{r}) - \phi_v(\boldsymbol{r}) \tag{2.43}$$

To obtain the reaction field, the Poisson equation is solved:

$$\nabla\left[\epsilon(\boldsymbol{r})\nabla\phi_s(\boldsymbol{r})\right] - \kappa(\boldsymbol{r})\phi_s(\boldsymbol{r}) = -4\pi\rho(\boldsymbol{r}) \tag{2.44}$$

Here $\epsilon(\boldsymbol{r})$ and $\bar{\kappa}(\boldsymbol{r})$ are the space-dependant dielectric constant and ion screening length. The charge distribution $\rho(\boldsymbol{r})$ is given by the point charges of the explicitly treated atoms using $\rho(\boldsymbol{r}) = \sum_A q_A \delta(\boldsymbol{r} - \boldsymbol{r_A})$. The vacuum electrostatic potential can easily be computed by setting $\epsilon(\boldsymbol{r}) = 1$ and $\bar{\kappa}(\boldsymbol{r}) = 0$ in Eq. 2.44.

As the solution of the differential equation Eq. 2.44 is too expensive to be used for a larger number of atomic configurations, a different approach is needed for molecular dynamics simulations. One possibility is the generalized solvent boundary potential (GSBP) approach [130].

In this approach, the system is partitioned into an inner region and an outer environment, where the dielectric property can vary e.g., containing implicitly treated parts of the protein, bulk solvent, a membrane slab, see Fig. 2.3. Atoms in the inner region are allowed to move during the simulation whereas atoms in the outer region are fixed.

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:

$$\Delta W_{elec} = \Delta W^{(oo)} + \Delta W^{(io)} + \Delta W^{(ii)} \tag{2.45}$$

The contribution from the outer-outer part can be computed analogously, but needs not to be taken into account as it is constant during the simulation. The contribution of the inner-outer part can be computed in a straightforward manner, as outlined above:

$$\Delta W^{(io)}(\boldsymbol{r}) = \sum_{A\in\text{inner}} q_A \phi_{rf}^{(o)}(\boldsymbol{r}_A) \tag{2.46}$$

where $\phi_{rf}^{o}(\boldsymbol{r}_A)$ is the reaction field due to the outer region atoms, which have 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 the *total* electrostatic potential due to the outer region atoms, $\phi_s^{(o)}$ is actually saved instead of the reaction field potential ($\phi_{rf}^{(o)}$) so that direct Coulombic interaction between the inner and outer region atoms is included. In principle also the inner-inner contribution could be computed in this way, however

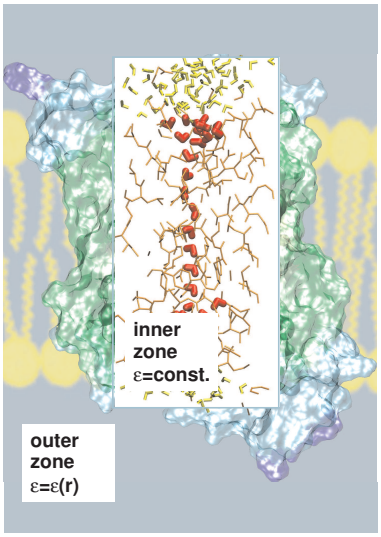


Figure 2.3: Example for electrostatic embedding using the generalized solvent boundary potential. The atoms in the inner zone are treated explicitly, while the protein atoms in the outer zone are treated through their electrostatic potential. The dielectric constant $\epsilon(r)$ in the outer zone is determined by the geometry of the protein, solvent and the membrane, if included.

as the charges move, the electrostatic potential would have to be recomputed in every single step of the simulation. Instead this part is expressed as

$$\Delta W^{ii}(\boldsymbol{r}) = \frac{1}{2}\iint d\boldsymbol{r}d\boldsymbol{r}'\rho^{(i)}(\boldsymbol{r})G(\boldsymbol{r},\boldsymbol{r}')\rho^{(i)}(\boldsymbol{r}') \tag{2.47}$$

where $G(\boldsymbol{r},\boldsymbol{r}')$ is the Green's function for the reaction field for composing the reaction field at $\boldsymbol{r}$ due to the charge at $\boldsymbol{r}'$:

$$\phi^{(i)}_{rf}(\boldsymbol{r}) = \int d\boldsymbol{r}'G(\boldsymbol{r},\boldsymbol{r}')\rho^{(i)}(\boldsymbol{r}') \tag{2.48}$$

This problem can be simplified to an algebraic problem if the charge distribution is expressed using a set of basis functions $b_m(\boldsymbol{r})$:

$$\rho^{(i)} = \sum_m c_m b_m(\boldsymbol{r}) \tag{2.49}$$

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

The inner-inner contribution to the reaction field free energy the takes the following form:

$$\Delta W^{ii}(\boldsymbol{r}) = \frac{1}{2}\sum_{mn} Q_m \left[\sum_{ij} O^{-1}_{im}M_{ij}O^{-1}_{jn}\right] Q_n = \frac{1}{2}\sum_{mn} Q_m M^0_{mn} Q_n \tag{2.50}$$

using:

$$M_{ij} = \iint d\boldsymbol{r}d\boldsymbol{r'} b_i(\boldsymbol{r}) G(\boldsymbol{r}, \boldsymbol{r'}) b_j(\boldsymbol{r'}) \tag{2.51}$$

generalized reaction field matrix

$\mathbf{M}^O$ is termed the *generalized reaction field matrix* and the numerical representation of the Green's function in the b_m basis. For a given geometry of the dielectric environment $\mathbf{M}^O$ needs to be computed only once via solving the Poisson-Boltzmann equation and does not depend on the instantaneous configuration of the nuclei. The choice of basis functions depends on the geometry of the problem. For globular proteins a spherical boundary often is the best choice and the basis functions are spherical harmonics [122, 130]. For membrane proteins in some cases, a rectangular boundary is more appropriate [124, 130]. Accordingly, Legendre polynomials are taken as the basis functions.

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

As described in a recent work by Schaefer et al. [122], the GSBP approach can be easily extended to a general QM/MM framework. For this purpose, it is assumed that all the QM atoms are within the inner zone.

The energy contributions to the QM/MM electrostatics from the reaction field then can be written as:

$$\Delta W^{(ii),QM/QM} + \Delta W^{(ii),QM/MM} + \Delta W^{(io),QM/MM} =$$
$$\frac{1}{2}\sum_{mn} Q_m^{QM} M_{mn} Q_n^{QM} + \frac{1}{2}\sum_{mn} Q_m^{QM} M_{mn} Q_n^{MM} + \int d\boldsymbol{r}\rho^{QM}(\boldsymbol{r})\phi_s^{(o)}(\boldsymbol{r}) \tag{2.52}$$

The elements of the vector Q^{MM} differ from the full charge vector by the fact that modifications required at the QM/MM frontier when cutting bonds (*vide infra*) are accounted for [122]:

$$Q_n^{MM} = \sum_{A\in\text{inner,MM}} \left(q_A + q_A^{shift}\right) b_n(\boldsymbol{r}_A) \tag{2.53}$$

Here 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 (chapter 4). Given the paradigm of point charges of the MM atoms acting on the Mulliken point charges of the QM atom ins the current implementation of the SCC-DFTB/CHARMM interface, the elements of the vector Q^{QM} are computed using:

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

The matrix elements $\widehat{H}_{\mu\nu}$ of the Hamilton operator are augmented by the following term due to the contributions from the reaction field:

$$\widehat{H}_{\mu\nu}^{GSBP} = \frac{1}{2} S_{\mu\nu} \sum_{B\in\text{QM}} \left[\Gamma_{CB}(\boldsymbol{R_C}, \boldsymbol{R_B}) + \Gamma_{DB}(\boldsymbol{R_D}, \boldsymbol{R_B})\right] \Delta q_B + \frac{1}{2} S_{\mu\nu} \left[\Omega(\boldsymbol{R_C}) + \Omega(\boldsymbol{R_D})\right]$$
$$\text{with } \mu \in C, \nu \in D \tag{2.54}$$

The basis functions μ, ν are located on the atoms C and D respectively and $S_{\mu\nu}$ is the respective overlap matrix element. Here, $\Gamma_{AB}(\boldsymbol{R_A}, \boldsymbol{R_B})$ describes the difference between the vacuum and reaction field interaction between QM atoms using the generalized reaction field matrix M^0:

$$\Gamma_{AB}(\boldsymbol{R_A}, \boldsymbol{R_B}) = \sum_{mn} b_m(\boldsymbol{R_A}) M^0_{mn} b_n(\boldsymbol{R_B}) \tag{2.55}$$

The reaction field due to MM atoms in the outer and inner zone is described using for a QM atom at $\boldsymbol{R_A}$ is described as:

$$\Omega(\boldsymbol{R_A}) = \sum_{mn} b_m(\boldsymbol{R_A}) M^0_{mn} Q^{MM}_n + \phi^{(o)}_s(\boldsymbol{R_A}) \tag{2.56}$$

The applicability of the SCC-DFTB/QMMM-GSBP framework has been demonstrated using the description of water structure within proteins [122, 124] and pK_a computations [123].

Chapter 3

Exploring the potential

3.1 Introduction

For obtaining data on molecules such as minimum-energy configurations, barriers to reactions, vibrational spectra, their potential has to be explored. Here, we will focus on methods for describing reactivity. There, we can distinguish between statical and dynamical methods [1].

Methods described here as static are focussed on locating stationary points, i.e., minima and saddle points[2] on the *potential energy surface (PES)* using *local* minimization.

potential energy surface (PES)

In contrast, dynamical methods explore the *free energy surface*. In this work, molecular dynamics is used for this purpose. Other powerful methods such as Monte-Carlo based schemes exist. Monte-Carlo based methods alter the configuration of the system by a set of pre-determined moves of atoms or groups of atoms. Given rules on acceptance or rejection of a given move, various ensembles can be simulated [131].

In the following, we will outline procedures mainly used in this work: Methods for obtaining minimum energy pathways and fundamentals of molecular dynamics and free energy computations.

Underlying techniques such as minimization will not be explained here, as a detailed review would be beyond the scope of this work. Instead the reader is referred to corresponding mathematical literature and textbooks in computational physics/chemistry for specific applications [91].

3.2 Minimum energy pathways

Minimal energy pathways (MEP) are paths on the PES connecting local minima, containing one or more saddle point(s) and having a vanishing gradient orthogonal to the tangent to the path in every point of the path (see Fig. 3.1). MEP are a valuable

Minimal energy pathways (MEP)

[1] In a mathematical sense, the methods described as statical below, such as the search for minima and minimal energy pathways can be described as dynamical processes. The distinction here is based on a physical point of view.

[2] more precise: saddle points of first oder

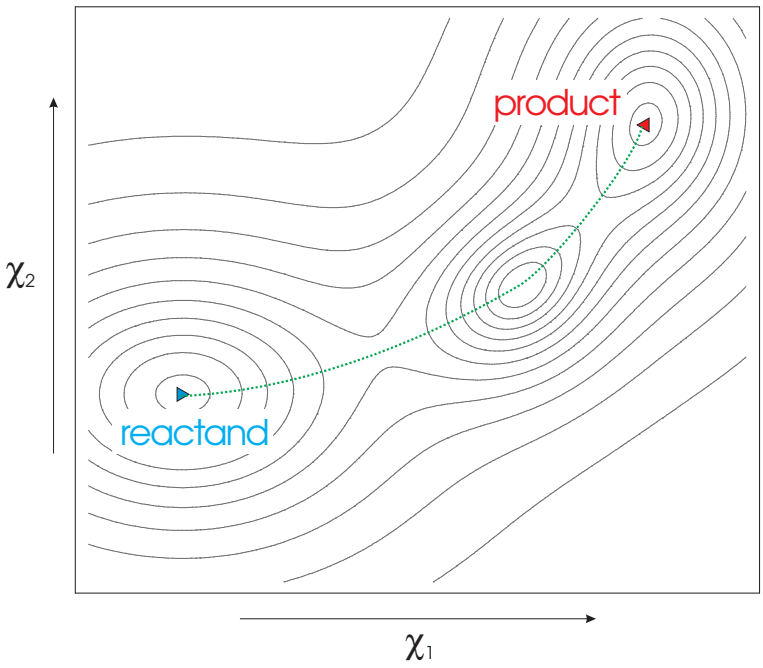


Figure 3.1: Schematic cross section through the potential energy surface. The energy is represented as a function of coordinates χ_1, χ_2. The minimum energy pathway connects the reactand and product wells.

tool for examining reactivity. The MEP needs not be unique, as multiple mechanisms can exist for realizing the reaction. However, not every MEP is realistic and realized in nature.

If the saddle points are known, the MEP can be obtained using steepest descent techniques. A number of strategies exist for locating saddle points [91, 132]. However, locating proper saddle points in high-dimensional systems can get involved. There, the predominant method for obtaining MEPs is to scan the system in the direction of a guess for the reaction coordinate. The obtained pathway is then refined using suitable algorithms. Ideally, the scanning step can be skipped and the guess for the minimum energy pathway can be obtained using interpolation between the reactand and product. However, in many cases this leaves the problem of choosing corresponding reactand and product states.

In the following, first a method for obtaining estimates of MEPs will be explained followed by methods of refinement. Further pitfalls of this methodology and the concept of MEPs will be illustrated.

3.2.1 Adiabatic mapping

adiabatic mapping

For obtaining an estimate of the potential energy surface for a reaction where just one or two coordinates are relevant, *adiabatic mapping* (AM), also referred to as coordinate driving, can be used [133]. The assumption behind this technique is that the reaction pathway can be obtained by gradually changing the reaction coordinate(s) and allowing all orthogonal degrees of freedom to relax in each step. The restric-

tion to a particular value of the reaction coordinate ζ can be imposed by a harmonic restraint:

$$E_{AM} = \frac{1}{2} k_\zeta \left(\zeta - \zeta_0\right)^2 \tag{3.1}$$

By altering the target value ζ_0 and minimizing[3] the system in each step, a scan of the PES is performed. The obtained energy maxima in the pathway are approximations to the saddle points for the reaction. The quality of this approximation depends apart from the step size used for scanning on how well the coordinate ζ describes the relevant reaction.

The task of picking a suitable reaction coordinate for obtaining meaningful guesses of the MEP can be challenging. For instance, for short range proton transfer, the geometrical asymmetric stretch coordinate is an obvious choice. Even though for more complex sequences of proton transfer linear combinations of asymmetric stretch coordinates can be useful, their use gets cumbersome for a large number of hops in the connecting wire. Details and an alternative suggestion will be given in Chap. 5.

3.2.2 Refinement

Adiabatic mapping can miss the transition state or underestimate the barrier significantly. This is especially true, if the *guess* for the reaction coordinate does not capture essential degrees of freedom of the actual reaction [134]. In this case, the coordinates of the system change discontinuously in the adiabatic mapping.

One method to refine the approximate pathway (e.g. obtained using adiabatic mapping) is *conjugate peak refinement (CPR)* [55]. CPR refines the path by converging any local maxima along the path to a saddle point on the potential energy surface. If the reaction path involves multiple transition states, each of them will have its saddle point included along the refined path. These saddle points are connected to each other by the steepest descent pathway that descend from them and form a continuous reaction coordinate between the stable reactant and product states. This continuous reaction coordinate represents the minimum energy pathway.

conjugate peak refinement (CPR)

Alternatives to the CPR algorithm are for instance the nudged elastic band (NEB) method suggested by Johannessón [54] and the self-penalty walk by Czerminski et al [135]. The disadvantage of these techniques is that the obtained pathway generally does not contain the saddle point(s), although an adaptive version of NEB exists which iteratively approaches the saddle point [136].

3.2.3 Caveats

One obvious limitation of the MEP approach is that the change in entropy between product, transition and reactand states is not taken into account. At finite temperature in systems with low barriers and a large conformational space, it may be more appropriate to explore the free energy surface.

[3] In the following the term "minimizing" will refer to minimizing the potential energy function with respect to the coordinates, as is customary in the literature.

A problem described in the literature is the problem of picking a suitable reactand state. This was recently illustrated in a simulation of by Klähn et al. studying the mechanism of phosphate cleavage in the Ras-GAP complex. In a molecular dynamics simulation of 1 ns, the potential energy of the reactand state was observed to fluctuate by about 30 kcal/mol [56]. The related fluctuations of the exothermicities ΔE and activation energy $E^{\ddagger}$ obtained from minimum energy pathways were not as pronounced but still were on the order of 10 and 5 kcal/mol respectively.

Another, closely related caveat is that the energetics of the pathway are biased by the starting point of the reaction path search. We can distinguish two partially competing effects:

- Overrelaxation
  Structures which are later in the pathway are potentially lower in energy than anticipated. The reason for this is that during the minimization degrees of freedom which are orthogonal to the reaction pathway can get unlocked and the system undergoes transitions which are not related to the reaction being studied.

- Underrelaxation
  Depending on the starting point the environment, in particular the solvent, favorizes the respective reactand.

The problem of overrelaxation can be overcome by performing sequentially adiabatic mapping forwards and backwards, i.e. in both directions of the reaction, until self-consistent pathways are found. A suitable reaction coordinate for a reaction should be able to describe both the forward and backward reaction.

The problem with underrelaxation can occur in reactions with large scale reorientation or significant charge shift. For instance in long range proton transfer, the solvent structure around both the donor and acceptor can be significantly different in the reactand and product state. Solvent reorientation is associated with barriers, which despite being low cannot be overcome in a local minimization based approach. As solvent structure determines proton affinities dramatically, the energetics of the reaction will be falsified. According to Marcus theory applied to electron transfer [137], reorientation of the solvent plays a crucial role.

The problems outlined here are not as pronounced in shorter proton transfer pathways or if the solvation structure around the reactive residues cannot change as dramatically.

Summarizing, the MEP provides a static description of the reaction. Though it can provide a useful qualitative understanding of the overall process, ideally a dynamic description of the process is necessary.

The energetics obtained computing MEPs for long-range proton transfer are only of qualitative nature. Appropriate sampling of the reactand state and computing a large number of minimum energy pathways could lead to a better understanding of the problem. Alternatively, using molecular dynamics to sample the system around the transition state and running minimum energy pathway searches from the snapshots may be a viable alternative for predictions of PT. Starting from the transition

state, the solvent structure will be optimal for proton transfer. Combined with averaging over a large number of pathways, this technique can provide quantitative information [138] using Marcus theory applied to proton transfer [137].

To overcome the problems of MEPs, molecular dynamics simulations can be used to predict a free-energy related measure and take into account the fluctuations and reorganization of protein and solvent.

3.3 Molecular Dynamics

To obtain statistical information on observables in a molecular system, dynamic methods can provide valuable information. The conformational space accessible at room temperature is comparatively large with low barriers separating a multitude of conformers. This limits the use of static methods such the ones outlined above. By not only taking into account the potential energy, but also the kinetic energy, low barriers can be overcome and the PES can be explored more extensively.

In the following, we will consider the dynamics of a system described by the Hamiltonian H, composed of kinetic $\mathcal{T}$ and potential $\mathcal{U}$ energy contributions:

$$\mathcal{H} = \mathcal{T} + \mathcal{U} = \sum_i \frac{p_i^2}{2m_i} + U \tag{3.2}$$

Here, p_i and m_i are the momentum and mass of the ith atom.

In general when investigating a molecular system we are interested in observables which be related to experiment. According to statistical mechanics, the average of an observable A can be obtained by calculating :

$$\langle A \rangle = \frac{\int d\boldsymbol{r}^N exp(-\beta \mathcal{U}(\boldsymbol{r}^N)) A(\boldsymbol{r}^N))}{\int d\boldsymbol{r}^N exp(-\beta \mathcal{U}(\boldsymbol{r}^N))} \tag{3.3}$$

In this equation $\beta = 1/(k_B T)$ is the thermal factor, defined by Boltzmann's constant k_B and the temperature T. Neither of the integrals involved here can efficiently be evaluated accurately, as it is far too costly to sample the entire phase space. Molecular dynamics gives us the mean of sampling the phase space, obtaining statistical information on observables. Ideally, the average of the observables converge quickly enough to reach ergodicity. However, this need not be the case, if the conformational space separated into areas with substantial barriers in between (see below).

To sample the phase space, molecular dynamics methods solve Newton's equation of motion:

$$m_A \frac{d^2 \boldsymbol{r}_A}{dt^2} = -\nabla \mathcal{U}(\boldsymbol{r}_1, ...\boldsymbol{r}_N) \qquad \text{for } 1 \leq A \leq N \tag{3.4}$$

Several different *integrators* have been suggested for solving (Eq. 3.4). For instance the Verlet integrator relies on a series expansion of the position vectors with time [131]. Other, more sophisticated integrators exist – with specific advantages and pitfalls [139]. integrators

The *timestep* used in such an integrator is determined by the shortest-timelength timestep

event to be integrated. The rule of thumb is to choose a timestep which is shorter by a factor of 10 than the eigenfrequency of the fastest motion. Typically, in atomistic simulations, these are X-H vibrations with a period of about $10^{-14}s$. The timestep is hence chosen to be $1\,fs$.

To increase the timestep of simulations, the distance of heteroatom-hydrogen bonds can be kept constant. This can be achieved by different algorithms such as SHAKE or RATTLE which introduce additional constraints in a Lagrangian formalism into Newton's equations of motion and solve the equations for the Lagrangian multipliers iteratively [140]. By eliminating X-H vibrations, the timestep for integration can be increased to $2\,fs$.

The process of integrating Newton's equations of motions is deterministic. However, the trajectory of the system in phase space depends critically on the initial conditions. Slightly different conditions lead to exploration of distinct regions of conformational space. This fact is known as *Lyapunov instability* [131]. Indeed, due to this rapid divergence of trajectories, it can be favorable to run multiple short simulations with slightly different initial conditions instead of one long trajectory [141].

Lyapunov instability

3.3.1 Ensembles

Using Newton's equations with conservative forces, energy is a constant of motion. This allows simulating a microcanonical ensemble (NVE). Canonical (NVT) and grand-canonical ensembles (μVT) imply constant temperature.

Canonical ensemble

According to classical thermodynamics the average kinetic energy $\langle T_i \rangle$ and speed v_i of a particle relate directly to the temperature of the system:

$$\langle \mathcal{T}_i \rangle = \left\langle \frac{1}{2} m_i v_i^2 \right\rangle = \frac{3}{2} k_B T \tag{3.5}$$

An intuitive attempt to maintain the temperature of the system is to compute the sum over the kinetic energies of all particles, relate it to the macroscopic temperature and rescale the velocities periodically. However, strictly speaking this does not lead to an ensemble with constant temperature, but instead to an isokinetic ensemble, i.e., constant kinetic energy.

Andersen suggested coupling to a heat bath using a description of stochastic collisions of randomly chosen atoms [142]. A more popular, deterministic approach are the extended Lagrangian methods introduced be Andersen and Nosé and improved by Hoover [143, 144]. Here, the heat reservoir is included using additional degrees of freedom in a Lagrangian approach, assigning the reservoir a fictitious mass Q. This parameter determines the strength of coupling between the reservoir and the system to be examined.

3.3.2 Langevin dynamics

For finite systems, where the amount of explicit solvent is limited, a different method of simulation can be applied. In Langevin dynamics, the dissipation of energy is directly described through the addition of dissipative forces to Newton's equation:

$$m_A \frac{d^2 \boldsymbol{r}_i}{dt^2} = -\nabla \mathcal{U}(\boldsymbol{r}_1, ...\boldsymbol{r}_N) - m_A \beta_A \frac{d\boldsymbol{r}_i}{dt} + R(t) \qquad \text{for } 1 \leq A \leq N \tag{3.6}$$

T_0 and T are the temperatures of the bath and the instantaneous temperature of the molecular system respectively (Eq. 3.5). Here, β_A is the frictional constant associated with atom A. Typically, a frictional constant is only assigned for non-hydrogen atoms. $R(t)$ is a random force due to thermal fluctuations of the environment.

In the frequently used stochastic boundary conditions setup for slab models [127], the dissipative forces act only on the outer shell of atoms with varying friction constants β. Energy flows in and out of the core region through direct contact with the outer shell, which in turn exchanges energy with the environment through dissipation.

3.3.3 Limitations of molecular dynamics simulations

Generally, even in force field based simulations, the timescale which can routinely be sampled is on the order of nanoseconds. This is significantly shorter than the timescale of many important biological processes. Many biological processes take place on a much longer timescale. For instance, many conformational changes take place on a timescale of nano- to microseconds; the proton transfer studied in Chap. 6 takes place on the timescale of milliseconds.

Different strategies to overcome the gap of the timescales of molecular dynamics and reactive events, different techniques have been developed, including, e.g., conformational flooding [145], hyperdynamics [146] and umbrella sampling [147]. The later will be the subject of the next section, after an introduction to the concept of the potential of mean force.

3.4 Potential of mean force

The derivation for the potential of mean force is closely coupled to radial distribution functions [148]. Here a brief sketch will be given.

Let us consider a system consisting of *indistinguishable* N particles with a mass m. The (conservative) forces acting between the particles are described by the interaction potential $\mathcal{U}$:

$$H = \sum_i \frac{p_i^2}{2m_i} + \mathcal{U}(\boldsymbol{r}^N) \tag{3.7}$$

Let $P^{(N)}(\boldsymbol{r}_1, \boldsymbol{r}_2, ..., \boldsymbol{r}_N)$ be the *configurational distribution function*, which describes the

configurational distribution function

probability of a particular configuration $\{\boldsymbol{r}^N\}$, i.e., of finding particle number 1 at $\boldsymbol{r}_1$, particle number 2 at $\boldsymbol{r}_2$ etc.:

$$P^{(N)}(\boldsymbol{r}_1, \boldsymbol{r}_2, ..., \boldsymbol{r}_N) = \frac{1}{Z} \exp(-\beta \mathcal{U}(\boldsymbol{r}^N)) \qquad \text{with } Z = \int d\boldsymbol{r}^N \exp(-\beta U(\boldsymbol{r}^N) \tag{3.8}$$

In a system, where the particles cannot be distinguished, the particle density is more meaningful. It is obtained from the configurational distribution function by factoring in possible permutations of indistinguishable particles:

$$\rho^{(N)}(\boldsymbol{r}_1, \boldsymbol{r}_2, ..., \boldsymbol{r}_N) = N! \frac{1}{Z} \exp(-\beta \mathcal{U}(\boldsymbol{r}^N)) \tag{3.9}$$

reduced distribution function

The *reduced distribution function* describes the probability of finding one particle at $\boldsymbol{r}_1$, one particle at $\boldsymbol{r}_2$ ... and one particle at $\boldsymbol{r}_n$. We can obtain it from the particle density by integrating over all (N-n) remaining degrees of freedom:

$$\rho^{(n)}(\boldsymbol{r}_1, \boldsymbol{r}_2, ..., \boldsymbol{r}_n) = \int \rho^{(N)}(\boldsymbol{r}_1, \boldsymbol{r}_2, ..., \boldsymbol{r}_N) d\boldsymbol{r}_{n+1} d\boldsymbol{r}_{n+2} ... d\boldsymbol{r}_N \tag{3.10}$$

For $n = 1$ the reduced distribution function (Eq. 3.10) is equivalent to the particle density:

$$\rho^{(1)}(\boldsymbol{r}_1) = N/V = \rho \tag{3.11}$$

To normalize the description of interactions the dimensionless reduced distribution function $g^{(n)}$ may be introduced:

$$g^{(n)}(\boldsymbol{r}_1, \boldsymbol{r}_2, ..., \boldsymbol{r}_n) = \rho^{(n)}(\boldsymbol{r}_1, \boldsymbol{r}_2, ..., \boldsymbol{r}_n)/\rho^n \tag{3.12}$$

radial distribution function

For $n = 2$ the function introduced in (3.4) is termed the *radial distribution function* or pair correlation function. It is often written in the form $g(r)$ where $r = |\boldsymbol{r}_1 - \boldsymbol{r}_2|$. From equation we can deduce for this case:

$$g(r) = g^{(2)}(\boldsymbol{r}_1, \boldsymbol{r}_2) / \left(g^{(1)}(\boldsymbol{r}_1)\right)^2 \tag{3.13}$$

In order to describe the interaction of two particles among many other particles in a system we introduce an averaged potential $w^{(n)}(\boldsymbol{r}_1, \boldsymbol{r}_2, ...\boldsymbol{r}_n)$. To see the relevance of this potential, we first compute the average force on particle 1 for a fixed distance $\boldsymbol{r}_1 - \boldsymbol{r}_2$:

$$\begin{aligned}
&-\left\langle \frac{d}{d\boldsymbol{r}_1} \mathcal{U} \right\rangle_{\boldsymbol{r}_1, \boldsymbol{r}_2 \text{ fixed}} = \frac{-\int d\boldsymbol{r}_3, ..., d\boldsymbol{r}_N (d\mathcal{U}/d\boldsymbol{r}_1) e^{-\beta \mathcal{U}}}{\int d\boldsymbol{r}_3, ..., d\boldsymbol{r}_N e^{-\beta \mathcal{U}}} = \\
&= k_B T \frac{(d/d\boldsymbol{r}_1) \int d\boldsymbol{r}_3, ..., d\boldsymbol{r}_N e^{-\beta \mathcal{U}}}{\int d\boldsymbol{r}_3, ..., d\boldsymbol{r}_N e^{-\beta \mathcal{U}}} = k_B T \frac{d}{d\boldsymbol{r}_1} \ln \int d\boldsymbol{r}_3, ..., d\boldsymbol{r}_N e^{-\beta \mathcal{U}} \\
&= k_B T \frac{d}{d\boldsymbol{r}_1} \ln \int d\boldsymbol{r}_3, ..., d\boldsymbol{r}_N e^{-\beta \mathcal{U}} + k_B T \underbrace{\frac{d}{d\boldsymbol{r}_1} \ln \frac{N(N-1)}{Z}}_{0} = k_B T \frac{d}{d\boldsymbol{r}_1} \ln g(\boldsymbol{r}_1, \boldsymbol{r}_2)
\end{aligned} \tag{3.14}$$

The potential corresponding to the averaged force between two particles can hence be directly related to the radial distribution function. It is the reversible work, related to bringing two particles from infinite separation to a separating distance $r = |\boldsymbol{r}_1, \boldsymbol{r}_2|$. The *reversible work theorem* states:

reversible work theorem

$$w(r) = -k_B T \ln g(r) + c \tag{3.15}$$

This potential $w(r)$ is termed the *potential of mean force (PMF)*.

potential of mean force (PMF)

More generally, for a set of distinguishable particles, the reversible work theorem (Eq. 3.15) can easily be extended. The change here only introduces a constant multiplicative factor to $g(r)$.

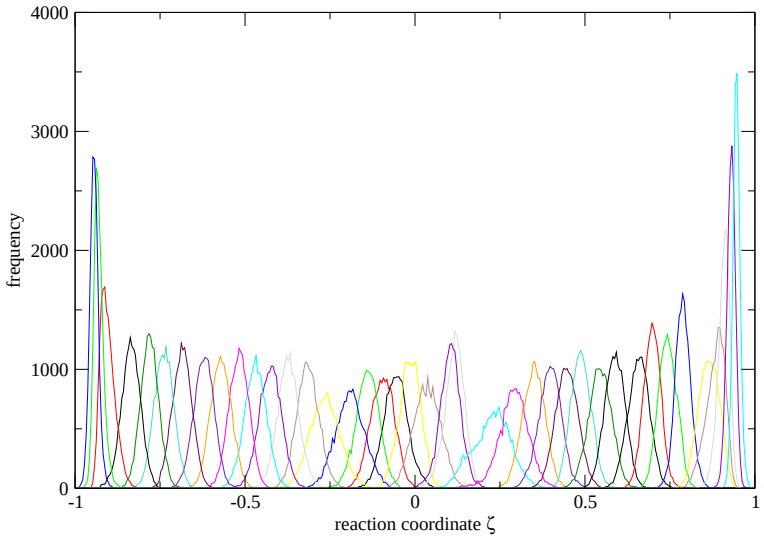


Figure 3.2: Overlapping histograms from umbrella sampling simulations.

Let us assume that we cannot sample a molecular system described by the potential energy function $\mathcal{U}(\boldsymbol{r}^N)$ accurately. Instead we are able to describe a system with a modified potential energy function $\mathcal{U}'(\boldsymbol{r}^N)$.

$$\mathcal{U}'(\boldsymbol{r}^N) = \mathcal{U}(\boldsymbol{r}^N) + \mathcal{U}_i(\zeta) \tag{3.16}$$

Here, $\mathcal{U}_i(\zeta)$ is a biasing potential and ζ is a reaction coordinate which describes the transition of the system between different states. For a simulation including $\mathcal{U}_i$ as the biasing potential, we can generate a histogram for the densities $\rho_i^*(\zeta)$ for the reaction coordinate ζ. Then we can obtain the potential of mean force $w_i(\zeta)$ for the non-biased system using [147]:

$$w_i(\zeta) = -k_B T \ln \left(\rho_i^*(\zeta)\right) + c_i \tag{3.17}$$

Generally, simulation with one particular umbrella potential can only sample part of the reaction. Simulations for one particular biasing potential are termed *windows*.

windows

Often, a large number of windows have to be computed for sampling the entire range of the reaction. The result of such a series of windows is a set of histograms, as shown in Fig. 3.2.

To obtain the PMF over the entire reaction coordinate from these histograms, the PMFs $w_i(\zeta)$ from each of the windows have to be connected. For overlapping PMFs, the constant offset c_i for each window can be obtained by matching the values at the interfaces between windows. An alternative method is to use the *weighted histogram analysis method (WHAM)* [149]. In both cases, a good overlap of the individual histograms is required to avoid ambiguity at the interfaces.

weighted histogram analysis method (WHAM)

In this work, the potential of mean force is computed using the WHAM equations.

Chapter 4

Evaluation of different QM/MM frontier treatments

4.1 Introduction

With quantum mechanical/molecular mechanical partitioning, 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 artifactual 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 involves inserting a *link atom*, which is typically chosen as a hydrogen atom, between the QM host atom (QMHA, see Fig. 4.1 and Tab. 4.1) 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 (however, see below) but not through van der Waals terms.

link atom

Instead of regular hydrogen atoms, hydrogen-like atoms or pseudohalogens have been used to terminate the QM region. In those approaches, the link atom coincides with the MMHA. The electronic nature of this atom is modified to mimic the behavior of the MM host atom or MM host group (MMHG). For example, in an approach described by Zhang et al. [150] for *ab initio* or DFT QM/MM, the boundary atom is described as a pseudohalogen with seven electrons, whose properties are adjusted to that of the atom it replaces using pseudopotentials. Antes et al. [151] suggested an alternative for semi-empirical QM/MM methods in which the integrals of the hydrogen-like boundary atom are re-parametrized to mimic the behavior of the host (e.g., methyl) group.

Table 4.1: Notations for atoms and groups used in QM/MM frontier treatments

MM atom	MMA	generic atom in the molecular mechanics (MM) partition
QM atom	QMA	generic atom in the quantum mechanics (QM) partition
MM host atom	MMHA	atom on the MM side of a severed bond at the frontier (Fig. 4.1)
MM host group	MMHG	charge neutral group including the MMHA
QM host atom	QMHA	atom on the QM side of a severed bond at the frontier (Fig. 4.1)
QM link atom	QML	link atom introduced between the QMHA and MMHA to cap the QM region, treated using QM (Fig. 4.1)
MM link atom	MML	link atom introduced between the MMHA and QMHA to balance dipole moment in the DLA approach, treated using MM (Fig. 4.1)

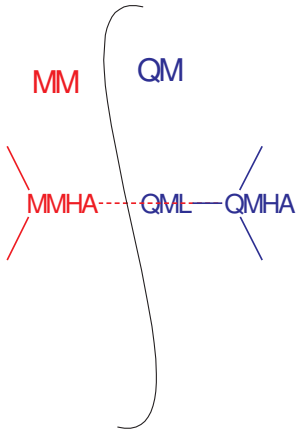


Figure 4.1: QM/MM partitioning across a covalent bond showing the MM host atom (MMHA), QM host atom (QMHA) and QM link atom (QML).

Other approaches to the QM capping problem include the *localized self consistent field (LSCF)* [152–155] and the *generalized hybrid orbital (GHO)* method [156], which use localized orbitals to describe the bonding at the QM/MM interface. Several studies using the LSCF approach in combination with DFT or HF [157–159] clearly demonstrated the capability and the pitfalls of the LSCF approach using small gas phase models [157, 158] as well as protein models [158, 159]. Moreover, Reuter et al. [160] showed in their systematic analysis using gas phase systems that the link atom approach and the LSCF method for AM1/CHARMM give comparable results. For small QM zones and MM atoms with large charges at the interface, the LSCF method gives worse results than the link atom approach. However, relatively large errors with both methods suggested that the QM/MM frontier should be chosen such that classical frontier atoms have small charges.

localized self consistent field (LSCF)

generalized hybrid orbital (GHO)

Within the link atom framework, it was argued that the treatment of QM/MM electrostatic interactions involving frontier MM atoms can influence the result substantially. Early implementation of the link atom approach sought to resolve the issue of spurious electrostatic interactions at the frontier by not taking into account interactions between the QM link atom and MM charges [102, 161]. Multiple authors showed [160, 162, 163] that this treatment suffers from erroneous polarization and other artifacts, due to distortions in the QM charge distribution as the frontier QM atom and the QM link atom experience significantly different electric fields. These effects are evident in both the electronic structure (e.g., dipole moment) and the energetics of protonation/deprotonation of the QM region.

A popular scheme, which is the default in CHARMM, excludes electrostatic interactions between the MMHA and all QM atoms (including the link atom) [162, 164–166]. This has been termed as the "*single link atom (SLA)*" treatment in a recent overview by Das et al. [115]. Noting that the SLA treatment leaves a fractional charge near the QM region, which can cause serious errors [115]. An alternative, which is termed the *excluded group (EXGR)* scheme in CHARMM and here, is to exclude QM/MM electrostatic interactions involving the entire MM host group (MMHG, see Fig. 2) [167]. Sherwood et al. suggested yet another variation, referred to as the *charge shift (CHSH)* scheme [168, 169], in which the charge of the MMHA is distributed to the neighboring atoms in the same group. Dipoles are placed on each of these atoms oriented in direction of the bond to the MMHA to compensate for the charge shift. The magnitude and position of these dipoles can be adjusted to obtain an accurate reproduction of the electrostatic potential in the QM region.

single link atom (SLA)

excluded group (EXGR)

charge shift (CHSH)

A number of authors suggested that no charges need to be exempt if the MM point charges are replaced by a *distributed ("blurred") Gaussian charge density (DG)* scheme - [115, 154, 170]. Das et al. [115] further suggested that, to counter the dipole moment created by the QM link atom, another link atom should be added to the MMHA. This link atom is termed the molecular mechanics link atom (MML) and the scheme is referred to as the *double link atom (DLA)* scheme. Impressive accuracy for proton affinities and deprotonation energies were obtained when DLA and the Gaussian blurring approaches were combined.[115] A remaining issue, however, is that the blurring width may not be generally transferable, e.g., different blurring widths were used for optimal protonation and deprotonation energies [115], which is problematic for studying proton transfer reactions.

distributed Gaussian charge density (DG)

double link atom (DLA)

Despite those previous efforts, the quantitative performance of various QM/MM frontier treatments in realistic condensed phase simulations is not well established. Previous studies have mainly used gas phase molecules as benchmark systems, and many such molecules are aliphatic compounds where little polarization is expected from the MM part. Only few studies have looked into the performance of different schemes for biologically relevant molecules [106, 160, 163, 170].

Here, performance of different link atom approaches available in CHARMM was systematically evaluated. In addition to typical gas phase test cases, a number of polar molecules where polarization is important were studied. Further, a number of enzyme systems were also included. The focus here is on QM/MM calculations using SCC-DFTB as the QM level. An assessment of different link atom schemes in the SCC-DFTB/MM framework will further help establish a generally effective QM/MM simulation protocol. The insights gained from the current study can also shed light into similar issues concerning QM/MM frontier treatments involving other QM methods, such as *ab initio* or more sophisticated DFT.

In Sec. 4.2, the QM/MM formulation for SCC-DFTB/MM is described and the different link atom schemes investigated here are clearly defined. In Sec. 4.3, results obtained using the various schemes are presented and results obtained using the various schemes are analyzed. The test cases include a number of polar and non-polar molecules in gas phase (Sec. 4.3.1). Examples for proton transfer reactions in both small molecules and enzymes will be given in Sec. 4.3.2. The chapter concludes in Sec. 4.4.

4.2 Methods

In this section, we first describe different descriptions of the QM/MM electrostatic interaction. This is followed by the definition of different link atom schemes studied here. Finally, computational details for all test calculations are described.

4.2.1 Description of QM/MM electrostatics

In the commonly available implementation of the QM/MM interface between SCC-DFTB and CHARMM, the electrostatic interaction between the MM zone and the QM zone is described through point charges on the MM atoms [114]. The expression for the Hamiltonian in this case was given in Eq. 2.40.

To achieve a balanced treatment of electrostatics between MM atoms and QM atoms at different distance separation, a number of authors [115, 154, 170] suggested using blurred Gaussian charge distributions instead of point charges for the MM atoms in evaluating QM/MM interactions. The blurred charge distribution ρ for a MM atom with total partial charge Q_A located at $\boldsymbol{R}$ is given as

$$\rho(\boldsymbol{r}) = Q_A \left(\frac{1}{\sigma\sqrt{2\pi}} \right)^{3/2} e^{-(\boldsymbol{r}-\boldsymbol{R})/2\sigma^2} \tag{4.1}$$

where σ is termed the blurring width. The corresponding contribution to the Hamil-

tonian operator similar to Eq. 2.40 is:

$$\widehat{H}^{QM/MM}_{elec,blurred} = \sum_{A\in MM}\sum_{B\in QM}\frac{Q_A\,\Delta q_B}{|\boldsymbol{r}_A-\boldsymbol{r}_B|}\mathrm{erf}\left(\frac{|\boldsymbol{r}_A-\boldsymbol{r}_B|}{\sigma}\right) \tag{4.2}$$

Finally, in the charge shifting model (CHSH) [168, 169], the QM atoms also interact with dipoles added to the MM atoms in the frontier region. The corresponding contribution to the Hamiltonian from a dipole $\boldsymbol{p}$ centered at $\boldsymbol{R}$ is:

$$\widehat{H}^{QM/MM}_{elec,dipole} = \sum_{B\in QM}\Delta q_B\frac{\boldsymbol{p}\cdot(\boldsymbol{r}_B-\boldsymbol{R})}{|\boldsymbol{r}_B-\boldsymbol{R}|^3} \tag{4.3}$$

4.2.2 Different link atom schemes

In all test calculations, the QM partition is capped by hydrogen link atoms. After the work was essentially finished, Gao et al. [171] implemented the GHO approach for SCC-DFTB/MM in CHARMM; therefore, the GHO results have not been compared to the link atom based approaches here. As described in Sec. 4.1, different link atom schemes tested here differ in the way that electrostatic interactions at the QM/MM interface are treated:

- *Single Link Atom (SLA)* [162, 164–166]:
  MM partial charges are represented as point charges; charge on the MM host atom (MMHA) is excluded from QM/MM electrostatics.
- *Excluded Group (EXGR)* [167]:
  MM partial charges are represented as point charges; charges on the entire MM host group (MMHG, see Fig. 2) are excluded from QM/MM electrostatics.
- *Distributed Gaussian (DG)* [115, 154, 170]:
  MM partial charges are represented as Gaussian charge distributions centered on the respective atoms; no exclusions for QM/MM electrostatic interactions are made. Identical blurring widths are used for all MM atoms.
- *Double Link Atom (DLA)* [115]:
  MM partial charges are represented as Gaussian charge distributions centered on the respective atoms; no exclusions for QM/MM electrostatic interactions are made. In addition to the QM link atom, a MM link atom (MML) is added to the MMHA and placed along the bond between the MMHA and the QMHA. The MML bears a partial charge polarizing all QM atoms, and the magnitude of its charge is subtracted from the MMHA to maintain charge neutrality.
- *Divided (DIV) frontier charge*:
  MM partial charges are represented as point charges. The charge of the MMHA is deleted and distributed evenly over the remaining MMHG. Similar to CHSH, but no dipoles are added to the MMHG atoms.
- *Charge Shift (CHSH)* [168, 169]:
  Redistribution of charges as described for the DIV scheme. A dipole is added

Table 4.2: Terms and interactions considered for various QM/MM schemes

Bonded terms		
	bond	MMA-X
	angle	X-MMA-X
	dihedral	X-MMA-X-X or X-X-MMA-X
	improper	MMA-X-X-X (first atom is central atom)
Nonbonded terms		
QMA-MMA	electrostatics	interaction through eqs (2.40),(4.2),(4.3)
	van der Waals	computed classically
QMA-QMA	electrostatics	not computed classically
	van der Waals	not computed
Nonbonded terms involving link atoms		
QML	electrostatics	interaction through eqs (2.40),(4.2),(4.3)
	van der Waals	invisible to both MMA and QMA
MML	electrostatics	classical interactions with all MMA interaction with all QMA through Eqs. (2.40),(4.2),(4.3)
	van der Waals	invisible to both MMA and QMA

X is either MMA or QMA

at the location of each of the remaining MMHG atoms to compensate for the charge shift from the MMHA. This dipole is exactly opposite to the dipole introduced by the amount of shifted partial charge from the MMHA to the respective atom multiplied by the distance between them.

Note that in almost all cases where partial charges are modified (SLA, EXGR, DIV, CHSH), such modifications are made *only* for the QM/MM interactions. This is crucial as otherwise the accuracy of the carefully parametrized MM/MM interaction by empirical force fields would be compromised. The only exception in which classical interactions are actually modified is the DLA approach. Furthermore, the sum of MM charges interacting with the QM atoms is conserved in all modifications except for SLA methods and EXGR when the overall charge of the MM host group is not neutral (no such cases are presented here).

Although the link-atom based frontier treatments discussed above were used only in the CHARMM framework here, these schemes are applicable in the context of other MM force fields. For schemes that involve modifying charges on the MM host group (CHSH [168, 169] and EXGR [167], DIV), it is not necessary that the host group has an integer charge. In fact, the CHSH and EXGR schemes were initially developed to modify charges within a given number of bonds from the QM or MM host atom; the DIV scheme can be used in the same spirit.

In addition to electrostatics, other terms also have to be considered for QM/MM interactions. Table 4.2 summarizes terms included and discarded in QM/MM calculations as implemented in SCC-DFTB/CHARMM [114] with different link atom schemes.

4.2.3 Computational details

All computations were carried out using the CHARMM27 forcefield [58] and the SCC-DFTB method [59]. The QM/MM interface implemented by Cui et al. [114] into the CHARMM program [61] was modified to include the DIV and CHSH approaches as well as the DG and DLA schemes described by Das et al. [115]. No electrostatic cutoffs were used for the gas phase models. No QM/MM restraints were used except for the DLA approach, in which the distance between the MM link atom (MML) and MMHA was restrained to be 1.1 Å and the MML was restrained to be along the bond connecting MMHA and QMHA using a force constant of 4000 kcal/(rad^2·mol).

The group definitions in CHARMM, which are relevant in CHSH, DIV and EXGR schemes, are as follows: for aliphatic compounds, each methyl- and methylene-group is charge neutral and constitutes a group in terms of the MMHG. For biomolecules tested here, the MMHG is schematically displayed in Fig. 4.2 for amino acids and nucleotides.

For the gas phase systems, the energy of all molecules was minimized with respect to their coordinates at the respective QM or QM/MM levels thus all comparisons were made with adiabatic energetics. When computing proton affinities and deprotonation energies, it should be noted that the energy of a proton in SCC-DFTB is not zero; a value of 141.8 kcal/mol is used as in standard parametrizations. This energy term is not to be confused with the heat of formation of a proton. Rather, it is a consequence of the definition of the repulsive energy term in the SCC-DFTB formalism. A detailed explaination can be found in [63].

To obtain the rotational profiles, an adiabatic mapping along the relevant dihedral angle was made between 0 ° (syn) and 180 ° (anti) in steps of 5 ° . At each step, the geometry was fully optimized while the relevant dihedral angle was harmonically restrained with a force constant of 2000 kcal/(rad^2·mol).

For proton transfer, adiabatic mappings were performed using the anti-symmetric stretch coordinate involving the proton donor, the proton and the proton acceptor for obtaining initial pathways, which were further refined using the CPR algorithm [55] to locate the saddle points more precisely. The two enzyme systems studied here, Triosephosphate Isomerase (TIM) and Methyl Glyoxal Synthase (MGS), were taken from previous SCC-DFTB/MM studies using the stochastic boundary condition [127]. The systems contain approximately $\sim$ 4000 enzyme atoms and 1179 water molecules for TIM and $\sim$ 7700 enzyme atoms and 384 water molecules for MGS. The QM region contains the substrate (DHAP) and side chains of two essential catalytic residues (His95 and Glu165 in TIM, His98 and Asp71 in MGS). For more details, refer to previous publications [28, 64, 172, 173]. In the enzyme systems electrostatic force shifting with a cutoff of 13.0 Å was used for MM/MM interactions. No cutoffs were used for QM/MM interactions in any of the models.

Finally, because the focus here is on the influence of QM/MM frontier treatment, full SCC-DFTB results were used as the reference for all test calculations on small molecules. For the gas phase models no cutoffs where used.

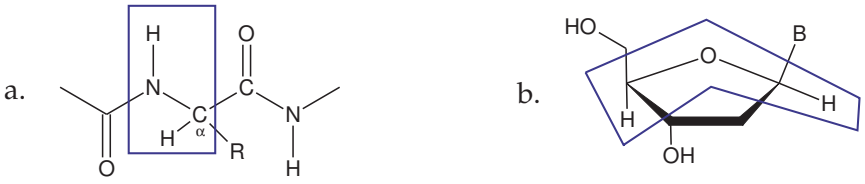


Figure 4.2: Typical MM host group (MMHG) for amino acids and proteins (a) and deoxyribose (b) in the framework of CHARMM force field. R substitutes the side chains. B substitutes the purine or pyrimidine base.

4.3 Results and Discussion

In the following different link atom approaches using a number of model systems and model reactions are compared. In Sec. 4.3.1 we explore the performance of different frontier treatments for small gas phase systems with different sizes of QM partition and MM partition. Specifically, we study proton affinities and deprotonation energies of alcohols and carbonic acids. We investigate how well energetics for different rotamers are reproduced using different link atom schemes.

To explore the situation in compounds with a more inhomogeneous electrostatic potential we examine protonation and deprotonation energies of amino acids and model DNA bases in Sec. 4.3.1.

Finally, to make the transition from absolute energies to reactions, we study proton transfer in an adenosine-thymidine Watson-Crick basepair, in two realistic enzyme systems, TIM and MGS in Sec. 4.3.2.

4.3.1 Results for gas phase models

Different sizes of MM region

A major assumption in QM/MM applications is that the electronic structure of the QM region is localized in nature and perturbations due to the environment can be described well with a MM treatment for the environment. As a first test of the stability of QM/MM frontier treatments, deprotonation energies for a series of aliphatic alcohols and carbonic acids were studied (see Tab. 4.3). In both cases we kept the size of the QM partition constant but varied the size of the molecule (and therefore the size of the MM partition). For the alcohols, a methanol fragment was considered QM for models A-D and an ethanol fragment was taken to be QM for models E-H. For the carbonic acids (models I-M), the QM partition includes the ethanoic acid fragment (see Tab. 4.3).

The general trend is that the EXGR and DLA/1.7Å give better results than all other schemes. The widely used SLA approach, which is default in CHARMM, gives the largest error on the order of 20 kcal/mol! The accuracy of DLA and DG schemes depend rather sensitively on the blurring width, especially with a small QM region; changing it from 1.7Å to 1.0Å increases the error by the order of 10 (2) kcal/mol when using the methanol (ethanol) QM fragment for the alcohols. As the size of the

QM region increases from methanol (A-D) to ethanol (E-H), the error in all link atom schemes decreases, as expected. The sensitivity on the blurring width in DG and DLA results is also reduced significantly. One interesting observation is that for all three classes of molecules considered in Tab. 4.3, the deviation from full SCC-DFTB results varies very little as a function of the MM size, which shows that the error to a large extent is due to approximate treatment of frontier QM/MM interactions. The only exception found is for EXGR, which gives much smaller errors when the MM region contains merely a methyl group. With EXGR, the partial charges on the entire frontier methyl/methylene group are excluded from QM/MM interactions; as a result, A/E/I have very similar deprotonation energies as B/F/J, respectively, all of which agree very well with full SCC-DFTB calculations. As soon as the MM partition is larger than a single group such that there are MM charges interacting with the QM region, the deviations increase in magnitude to a range comparable to DLA/1.7Å.

Although SCC-DFTB is used to describe the QM region, the qualitative trends are very similar to those found in previous studies using different QM methods. For example, models C and F were also studied by Reuter et al. [160] using an LSCF approach and the AM1 Hamiltonian. The errors compared to full QM calculations are 6.6 and 0.6 kcal/mol, respectively, which are close to the values of 6.5 and 1.5 kcal/mol found here with EXGR. For model G, Das et al. [115] found errors of -14.5, 4.5 and 3.2 kcal/mol for SLA, EXGR and DLA/1.7A, respectively, using B3LYP/6-31G* as the QM method; these values are very similar to the SCC-DFTB/MM results, which are -13.1, 4.5 and 4.7 kcal/mol, respectively. These observations support the hypothesis that the general trends observed here can be generalized to QM/MM approaches with other QM methods.

We note that none of the QM/MM approaches reproduce the trend of decreasing deprotonation energies with increasing molecule sizes for test cases in Tab. 4.3. This is not surprising since the effects due to non-polar MM groups are subtle and within the error bars of QM/MM calculations. For cases with more polar MM groups, QM/MM methods can indeed reproduce qualitative trends very well (see below).

Different sizes of QM region

An important choice in QM/MM simulations concerns the size of the QM region and therefore the separation between the QM/MM interface and the reactive site. Whether a small QM region is satisfactory depends critically on the accuracy of the MM force field and treatment of the QM/MM interface. To study the effect of the latter, deprotonation energies and proton affinities were computed for different link atom methods and varying QM sizes, using pentanol and pentanoic acid as test cases (see Tab. 4.4 and 4.5).

As shown in Tab. 4.4, the SLA approach gives the worst results for both pentanoic acid and pentanol with an average absolute deviation of $\sim$14 kcal/mol. Interestingly, the deviations do not decrease systematically with increasing size of the QM partition. For other link atom methods, by contrast, the error of QM/MM calculation does decrease monotonically with increasing size of the QM region. In the case of pentanol, for example, the error decreases from 6.6 (model Q) to 0.3 kcal/mol

Table 4.3: Deprotonation energies for QM fragments embedded in various MM environments (in kcal/mol)[a,b].

Model		SCC-DFTB	SLA	EXGR[c]	DLA/1.0Å	DLA/1.7Å	DG/1.0Å	DG/1.7Å	CHSH
A	**CH3-OH**	397.1	—	—	—	—	—	—	—
B	CH_3-**CH_2-OH**	397.0	369.2	397.6	417.8	405.2	415.0	407.6	410.7
	dev.		-27.8	0.6	20.8	8.2	18.0	10.6	13.7
C	CH_3-CH_2-**CH_2-OH**	395.5	379.0	402.0	414.8	405.2	412.4	407.7	410.2
	dev.		-16.5	6.5	19.3	9.7	16.9	12.2	14.7
D	CH_3-CH_2-CH_2-**CH_2-OH**	394.9	378.3	401.4	412.7	404.1	411.9	407.2	409.7
	dev.		-16.6	6.5	17.8	9.2	17.0	12.3	14.8
E	**CH_3-CH_2-OH**	397.0	—	—	—	—	—	—	—
F	CH_3-**CH_2-CH_2-OH**	395.5	371.9	397.0	402.5	399.4	404.8	402.0	403.1
	dev.		-23.6	1.5	7.0	3.9	9.3	6.5	7.6
G	CH_3-CH_2-**CH_2-CH_2-OH**	394.9	381.8	399.4	402.3	399.6	403.4	401.6	402.8
	dev.		-13.1	4.5	7.4	4.7	8.5	6.7	7.9
H	CH_3-CH_2-CH_2-**CH_2-CH_2-OH**	394.6	381.6	399.1	400.6	398.9	403.2	401.4	402.6
	dev.		-13.0	4.5	6.0	4.3	8.6	6.8	8.0
I	**CH_3-COOH**	366.4	—	—	—	—	—	—	—
J	CH_3-**CH_2-COOH**	365.7	338.8	366.3	373.3	367.7	375.5	372.0	374.0
	dev.		-26.9	0.6	7.6	2.0	9.8	6.3	8.3
K	CH_3-CH_2-**CH_2-COOH**	365.0	349.6	368.8	370.7	367.5	373.8	371.6	373.3
	dev.		-15.4	3.8	5.7	2.5	8.8	6.6	8.3
L	CH_3-CH_2-CH_2-**CH_2-COOH**	364.7	349.3	368.6	370.8	368.2	373.6	371.4	372.6
	dev.		-15.4	3.9	6.1	3.5	8.9	6.7	7.9
M	CH_3-CH_2-CH_2-CH_2-**CH_2-COOH**	364.5	349.9	368.3	370.8	368.1	373.6	371.2	372.4
	dev.		-14.6	3.8	6.3	3.6	9.1	6.7	7.9

a The QM-zone consists of the bold-faced portion of the molecule plus the link atom. Deviations are relative to the entire molecule computed using SCC-DFTB.

b The column SCC-DFTB indicates a calculation encompassing the entire molecule.

c For the molecules shown here, the DIV scheme is equivalent to the EXGR scheme and is hence not listed separately.

(model T) for EXGR, from 10.1 to 1.7 kcal/mol for DLA/1.7 Å, from 12.5 to 2.8 kcal/mol for DG/1.7 Å and from 15.0 to 3.4 kcal/mol for CHSH. We note that the best link atom results are only slightly worse than those reported for GHO/SCC-DFTB [171], which includes special parametrizations for the frontier atoms.

For the proton affinities (Tab. 4.5), similar trends are seen; e.g., the SLA approach shows a large mean error of 15 kcal/mol, while EXGR, DLA/1.7 Å, DG/1.7 Å and CHSH give much smaller deviations compared to full SCC-DFTB results. An interesting difference from deprotonation energies is that deviations in proton affinities are more or less constant with respect to the size of the QM region. This is likely due to the fact that the charge distribution in the cationic species involved in proton affinity calculations is highly localized in space, thus error in QM/MM calculations arises mainly from the frontier treatment and is independent from the size of the QM region. In deprotonation energy calculations, by contrast, anionic species are involved, which have a more delocalized charge distribution; therefore, QM/MM results are more sensitive to the size of the QM region. A generally poorer description of deprotonation energies in comparison to proton affinities was also observed by Reuter et al.[160].

For methods based on Gaussian blurring, i.e., DG and DLA, the results depend on the blurring width, as found in previous work and last subsection. Previous studies [115] found that different blurring widths are better suited for different properties; e.g., a blur width of 1.0 Å and 1.7 Å was found to give better deprotonation energy and proton affinity, respectively, with HF and B3LYP as the QM methods. The current calculations using SCC-DFTB as QM, in contrast, found that 1.7 Å gives better results for both quantities. This is confirmed by calculations with more extensive sampling of different blurring widths (Fig. 4.3). For both deprotonation energies and proton affinities, the sensitivity towards the blurring width decreases sharply with increasing size of the QM region, and this effect is less pronounced for proton affinity than deprotonation energy. While deviations in deprotonation energies decrease with increasing blurring width, errors may increase for proton affinities when the blurring width is beyond a certain value. The proton affinities of model Q and R are well described using a blurring width of 1.0 Å, and a blurring width of 1.7 Å seems to be a reasonable compromise for both proton affinity and deprotonation energy. We note that with a large blurring width the results should converge towards the EXGR values. Since EXGR may be problematic when there are important contributions from dipole and higher moments in the MMHG (see below), increasing the blur radius *ad libidum* is not generally robust.

Rotational profiles

A useful probe for the performance of the QM/MM frontier treatment is the energy profile for the rotation around the bond at the QM/MM interface. As shown in Eq. 2.38, this is described mainly via a classical MM dihedral term, although the QM/MM frontier interaction may also contribute. Since the MM dihedral terms were parametrized to be consistent with the MM electrostatics (e.g., some force fields including CHARMM include no or scaled 1-4 interactions [89]), unbalanced QM/MM frontier interaction may significantly perturb the rotational profile of the frontier bond. The test system chosen in this context is pentanol with different

Table 4.4: Deprotonation energies for varying QM zones (in kcal/mol)[a,b].

Model		SCC-DFTB	SLA	EXGR[c]	DLA/1.0Å	DLA/1.7Å	DG/1.0Å	DG/1.7Å	CHSH
Q	CH_3-CH_2-CH_2-CH_2-**CH_2-OH**	397.1	378.1	401.2	414.3	404.7	411.8	407.1	409.6
	dev.		-16.5	6.6	19.7	10.1	17.2	12.5	15.0
R	CH_3-CH_2-CH_2-**CH_2-CH_2-OH**	397.0	381.6	399.1	400.6	398.9	403.2	401.4	402.6
	dev.		-13.0	4.5	6.0	4.3	8.6	7.0	8.0
S	CH_3-CH_2-**CH_2-CH_2-CH_2-OH**	395.5	383.9	397.2	401.0	397.6	400.8	399.1	400.0
	dev.		-10.7	2.6	6.4	3.0	6.2	4.5	5.4
T	CH_3-**CH_2-CH_2-CH_2-CH_2-OH**	394.9	379.5	394.9	399.2	396.3	399.1	397.4	398.0
	dev.		-15.1	0.3	4.6	1.7	4.5	2.8	3.4
ref.	**CH_3-CH_2-CH_2-CH_2-CH_2-OH**	394.6	—	—	—	—	—	—	—
U	CH_3-CH_2-CH_2-**CH_2-COOH**	366.4	349.3	368.6	370.5	368.4	373.7	371.8	372.6
	dev.		-15.4	3.9	5.8	3.7	9.0	7.1	7.9
V	CH_3-CH_2-**CH_2-CH_2-COOH**	365.7	353.1	367.2	371.7	367.9	371.4	369.5	370.6
	dev.		-11.6	2.5	7.0	3.2	6.7	4.8	5.9
W	CH3-**CH_2-CH_2-CH_2-COOH**	365.0	348.5	364.9	369.7	366.5	369.6	367.6	368.3
	dev.		-16.2	0.2	5.0	1.8	4.9	2.9	3.6
ref.	**CH_3-CH_2-CH_2-CH_2-COOH**	364.7	—	—	—	—	—	—	—

a The QM-zone consists of the bold-faced portion of the molecule plus the link atom. Deviations are relative to the entire molecule computed using SCC-DFTB.

b The column SCC-DFTB indicates a calculation including only the capped QM fragment.

c For the molecules shown here, the DIV scheme is equivalent to the EXGR scheme and is hence not listed separately.

Table 4.5: Proton affinities for varying QM zones (in kcal/mol)[a,b].

Model		SCC-DFTB	SLA	EXGR[c]	DLA/1.0Å	DLA/1.7Å	DG/1.0Å	DG/1.7Å	CHSH
Q	CH_3-CH_2-CH_2-CH_2-**CH_2-OH**	-186.7	-169.9	-191.6	-198.8	-191.5	-198.1	-192.8	-195.0
	dev.		23.6	1.9	-5.3	2.0	-4.6	0.7	-1.5
R	CH_3-CH_2-CH_2-**CH_2-CH_2-OH**	-191.7	-179.8	-195.7	-203.0	-196.7	-199.7	-198.6	-198.0
	dev.		13.7	-2.2	-9.5	-3.2	-6.2	-5.1	-4.5
S	CH_3-CH_2-**CH_2-CH_2-CH_2-OH**	-194.6	-183.4	-196.1	-199.5	-196.4	-197.3	-197.7	-196.4
	dev.		10.1	-2.6	-6.0	-2.9	-3.8	-4.2	-2.9
T	CH_3-**CH_2-CH_2-CH_2-CH_2-OH**	-195.3	-180.6	-195.4	-199.9	-196.7	-197.5	-197.7	-196.2
	dev.		12.9	-1.9	-6.4	-3.2	-4.0	-4.2	-2.7
ref.	**CH_3-CH_2-CH_2-CH_2-CH_2-OH**	-193.5	—	—	—	—	—	—	—

a The QM-zone consists of the bold-faced portion of the molecule plus the link atom. Deviations are relative to the entire molecule computed using SCC-DFTB.

b The column SCC-DFTB indicates a calculation including only the capped QM fragment.

c For the molecules shown here, the DIV scheme is equivalent to the EXGR scheme and is hence not listed separately.

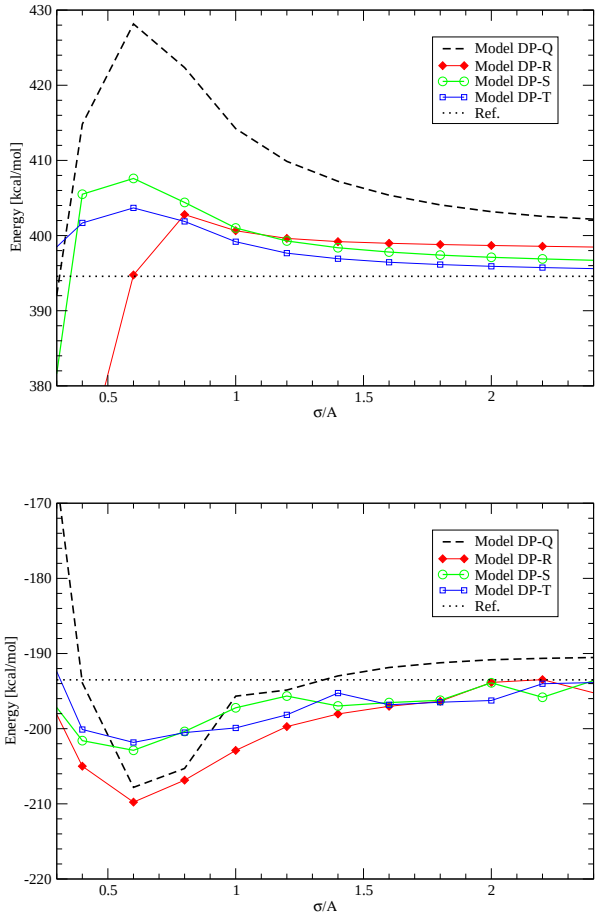


Figure 4.3: Deprotonation energies (a) and proton affinities (b) for the models introduced in Tab. 4.4 and 4.5 for varying Gaussian blurring width σ using the DLA scheme.

QM/MM partitions that correspond to models Q-S in Tab. 4.4; the rotational profiles are shown in Fig. 4.4 a-c, where all energies are relative to the *anti* conformer (180 °). In all cases full SCC-DFTB calculations yield too low rotational barriers compared to full MM results. This has also been observed in previous studies using semi-empirical treatments [174]. We thus focus on how close can various QM/MM treatments reproduce full MM results as in previous work [160].

Overall, except when the QM/MM frontier is close to the functional group (OH here), the deviations for different QM/MM approaches from the full MM rotational profile are small. For models R and S, the typical deviation of the various link atom schemes from full MM results is less than 0.5 and 1.3 kcal/mol, respectively. For model Q, in contrast, significant scattering for different link atom approaches are observed (Fig. 4.4 a). For example, while MM calculation predicts the *gauche* conformation to be 1.3 kcal/mol higher in energy than the *anti* conformer, the EXGR scheme finds the opposite trend. For the *syn* conformer, results of QM/MM calculations scatter to lower energies compared to full MM results. Here EXGR shows the largest deviation with an underestimation by 2.1 kcal/mol. The DLA/1.7Å is found to give results most similar to the full MM treatment.

The observed trend can be rationalized based on the charge distribution in the QM/MM interface region. In models R and S, atoms in the vicinity of the rotating bond have small partial charges and hence boundary QM/MM electrostatic interactions do not play an important role; as a result, the rotational profile is dominated by the classical MM dihedral term. In model Q, however, the rotating bond is close to the functional group (OH), which bears significant partial charges; therefore, frontier QM/MM electrostatic interactions may significantly modify the rotational profile, depending on the way such interactions are treated in different link atom schemes. This is most obvious for the EXGR scheme and model Q, where the *gauche* conformation is predicted to be more stable than the *anti* conformation. Still, this is the largest deviation for all cases shown here and the deviation with respect to full MM is smaller than 2 kcal/mol.

Dipole moments

Charge distribution in molecules is of crucial importance for reactivity and interaction with the environment. A reliable QM/MM scheme should hence be able to reproduce the charge distribution correctly, and one way to analyze this is to compare dipole moments computed with QM/MM against full QM results. Systems chosen here include ethane as an unpolar test case and ethanol/propanol as polar test cases.

As seen in Tab. 4.6, *all* QM/MM schemes yield non-vanishing dipole moments for ethane. The DLA scheme, which is motivated by obtaining correct dipole moments in QM/MM calculations[115], gives a large value (e.g., 1.0 Å) unless a large blurring width is used. For polar examples (ethanol and propanol) that involve a methanol QM region, the deviation from full SCC-DFTB result is small relative to the absolute value of the dipole moment for all link atom schemes, and the maximum deviation is only about 0.2 Debye. With a larger QM region in propanol, somewhat unexpectedly, very large errors (on the order of 1 Debye!) are observed for CHSH and

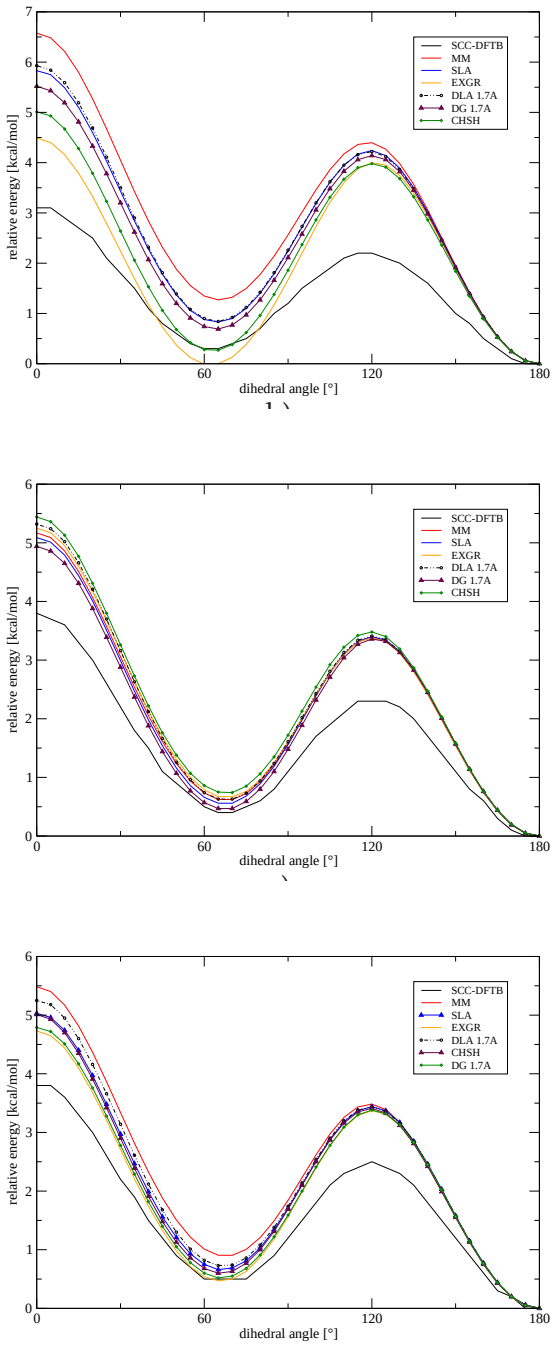


Figure 4.4: Rotational profile for (a) model Q of pentanol around the QM/MM frontier C-C bond: CH_3-CH_2-CH_2-CH_2-**CH_2-OH**, (b) model R of pentanol around the QM/MM frontier C-C bond: CH_3-CH_2-CH_2-**CH_2-CH_2-OH** and (c) model S of pentanol around the QM/MM frontier C-C bond: CH_3-CH_2-**CH_2-CH_2-CH_2-OH**. (QM zone in bold)

Table 4.6: Dipole moments (magnitude) for small molecules in Debye[a].

	CH_3-**CH_3**	CH_3-**CH_2OH**	CH_3-CH_2-**CH_2OH**	CH_3-**CH_2-CH_2OH**
SCC-DFTB	0.00	1.35	1.32	1.32
SLA	0.50	1.40	1.27	1.40
EXGR	0.50	1.19	1.16	1.66
CHSH	1.08	1.42	1.20	2.10
DG/1.0Å	1.59	1.14	1.32	2.56
DG/1.7Å	0.90	1.14	1.18	1.96
DG/2.5Å	0.62	1.16	1.17	1.73
DG/3.0Å	0.56	1.18	1.19	1.68
DLA/1.0Å	1.79	1.56	1.39	2.68
DLA/1.7Å	0.56	1.15	1.18	1.68
DLA/2.5Å	0.23	1.28	1.27	1.47
DLA/3.0Å	0.14	1.33	1.31	1.41

a The QM fragment is indicated in bold. b Science is supposed to be fun. If you read this you have obviously dug much deeper into my dissertation than most people. The first person to share the fun with me will receive a bottle of champagne.

DG/DLA calculations with a small blurring width. As the blurring width increases, both DLA and DG results improve substantially and the DLA converges towards the correct value faster. The fact that errors in the dipole moment increase for some link atom schemes with a larger QM region while errors in deprotonation energies decreases for most cases (Tab. 4.3) is not difficult to understand. Deprotonation energy is a differential quantity, thus certain inaccuracies of the electronic structure due to QM/MM frontier treatment can be masked due to error cancellation; dipole moment, on the other hand, is a more direct and sensitive probe of the electron distribution in *one* molecule. Being a collective variable, however, an accurate dipole moment *alone* does not necessarily imply an accurate treatment; a good example is that SLA gives rather decent dipole moments for the molecules tested here (Tab. 4.6) but produces large errors for deprotonation energies (Tab. 4.3).

Amino Acids

When studying reactions in proteins, one obvious way to partition the system is to take the side chains of interest into the QM partition and the rest into the MM partition, cutting across the C^{α}-C^{β} bond. It is important to establish robust schemes for such a partitioning. In contrast to the aliphatic alcohols and carbonic acids discussed above, amino acids show a non-uniform charge distribution in the backbone, thus the QM/MM frontier scheme is expected to have an even greater impact on the result compared to the cases studied above. The benchmark chosen involves deprotonation energies (Tab. 4.7) and proton affinities (Tab. 4.8) for the polar and charged amino acid side chains. These properties are of interest because change in protonation state is a major mechanism for amino acid side chains to be involved in general acid - general base catalysis [8]. In all calculations, the N-terminus was capped with an acetyl residue and the C-terminus with a methyl-ester, to mimic QM/MM partition inside a protein.

Overall, all models other than SLA and EXGR give rather similar performance. The SLA approach shows the largest deviations relative to full SCC-DFTB data with 16.2 kcal/mol for deprotonation and 19.2 kcal/mol for proton affinities. This is somewhat surprising (and alarming) as in many cases the protonation/deprotonation site is far from the QM/MM interface. For example, the deviation in the deprotonation energy for lysine is as high as 8.5 kcal/mol despite the large distance between the deprotonation site and the QM/MM interface. For tyrosine and histidine, significant deviations could be rationalized with unsatisfactory treatment for the polarization of the aromatic ring.

The EXGR results are mixed in quality. For deprotonation energies, the deviations for lysine are small and comparable with other methods. For the other amino acids, the deviations are about 10.6 kcal/mol and of similar magnitude as those for SLA. For the proton affinities, the deviations are generally lower (5.5 kcal/mol) and are dominated by the poor description of aspartate and glutamate.

The general difficulty associated with the EXGR approach is caused by the fact that the backbone N-H belongs to the same group as the $C^{\alpha}H^{\alpha}$ in the CHARMM force field, which means that a major dipole moment is excluded from QM/MM frontier treatment. This problem is alleviated to certain extent by the DG, DLA, DIV and CHSH schemes, which keep the N-H charges while still being careful about the charge neutrality at the interface (unlike SLA). Indeed, they give generally similar results, although the DIV approach appears to be most robust. For example, the DG and DLA calculations with a blurring width of 1.7 Å have an absolute average deviation of 2.6 and 1.5 kcal/mol, respectively, for the deprotonation energies. The CHSH and DIV models give average errors of 1.7 kcal/mol. However, only the DIV treatment gives a comparably reliable treatment for the protonation affinities (1.7 kcal/mol), while other methods produce much larger errors: the DG/1.7Å, DLA/1.7Å and CHSH methods have deviations of 4.0, 8.3 and 4.3 kcal/mol, respectively, and the DLA/1.0 Å scheme even had the problem that the aspartate anion is not stable and dissociated between C^{α} and C^{β}.

One interesting question at this point is whether the quality of the description can be improved by choosing a different QM/MM partitioning scheme. For instance, both the lysine and the glutamate side chains are sufficiently long for shifting the frontier bond to C^{β}-C^{γ}, which would put the QM/MM interface closer to the protonation/deprotonation site but leave the electrostatics of the backbone unperturbed (note that EXGR and DIV becomes identical with this partition). As seen in Tab. 4.8, the proton affinities of glutamate improve significantly except for the SLA approach which suffers from the decreased size of the QM zone. The error in CHSH did not improve either and became of different sign. The error for EXGR drops significantly to 0.4 kcal/mol, which is roughly the same as for the DIV scheme in the original C^{α}-C^{β} partitioning. For the deprotonation energies (Tab. 4.7), improvements in lysine are substantial for some approaches. The EXGR result improves from 3.7 to -2.2 kcal/mol, and both the DLA and the DG schemes as well as the CHSH approach showed similar improvements. For glutamate, a similar decrease in error was seen in the study of Reuter et al. using AM1 [160] as the QM method with different QM/MM partitioning schemes.

Table 4.7: Deprotonation energies for amino acids (in kcal/mol)[a,b] .

	SCC-DFTB	SLA	EXGR	DIV	DLA/1.0Å	DLA/1.7Å	DG/1.0Å	DG/1.7Å	CHSH
Ac-Lys-OCH_3	222.8	231.3	226.5	222.2	218.7	219.7	219.1	220.1	220.1
dev.		8.5	3.7	-0.6	-4.1	-3.1	-3.7	-2.7	-2.7
Ac-Lys-OCH_3 [c]	222.8	206.6	220.6	—	220.5	220.1	224.4	222.7	223.7
dev.		-16.2	-2.2	—	-2.3	-2.7	1.6	-0.1	0.9
Ac-His$^{\delta}$-OCH_3	375.1	390.7	385.8	376.0	370.2	371.4	368.6	372.2	372.1
dev.		15.6	10.7	0.9	-4.9	-3.7	-6.5	-2.9	-3.0
Ac-His$^{\epsilon}$-OCH_3	374.0	390.1	385.7	377.2	371.3	372.3	369.6	373.5	373.4
dev.		16.1	11.7	3.2	-2.7	-1.7	-4.4	-0.5	-0.6
Ac-Tyr-OCH_3	356.4	368.7	365.2	358.4	353.5	354.5	352.4	355.3	355.3
dev.		12.3	8.8	2.0	-2.9	-1.9	-4.0	-1.1	-1.1
rmsd		14.6	8.3	2.0	3.5	2.7	4.3	1.9	1.9

a Ac=acetyl protecting group,

b The column SCC-DFTB indicates the property of the entire molecule calculated with SCC-DFTB. For the remaining QM/MM calculations the side chain including the link atom is treated with SCC-DFTB while the backbone is treated using MM. Deviations are relative to the entire molecule computed using SCC-DFTB.

c QM/MM frontier between C^{β} and C^{γ}.

Table 4.8: Proton affinities for amino acids (in kcal/mol)[a,b].

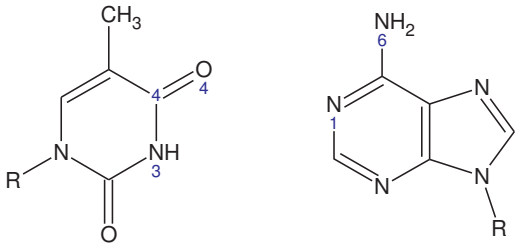
	SCC-DFTB	SLA	EXGR	DIV	DLA/1.0Å	DLA/1.7Å	DG/1.0Å	DG/1.7Å	CHSH
Ac-Asp-OCH_3	-356.7	-373.0	-368.5	-358.3	—	-350.4	-350.8	-353.2	-353.1
dev.		-16.3	-11.8	-1.6	—	6.3	5.9	3.4	3.6
Ac-Glu-OCH_3	-359.7	-371.4	-367.6	-359.6	-351.7	-354.2	-352.5	-356.1	-355.9
dev.		-11.7	-7.9	0.1	7.9	5.5	7.2	3.6	3.8
Ac-Glu-OCH_3 [c]	-359.7	-340.5	-359.3	—	-363.2	-355.8	-363.9	-361.7	-363.5
dev.		19.2	0.4	—	-3.5	3.9	-4.2	-2.0	-3.8
Ac-Tyr-OCH_3	-193.8	-200.7	-196.2	-191.3	-187.3	-187.8	-188.2	-188.9	-189.1
dev.		-7.0	-2.4	2.5	6.5	6.0	5.6	4.9	4.7
Ac-His$^{\delta}$-OCH_3 [d]	-255.1	-261.5	-252.7	-248.3	-243.3	-244.1	-243.2	-245.3	-245.0
dev.		-6.4	2.4	6.8	11.8	11.0	11.9	9.8	10.1
Ac-GluH-OCH_3	-205.2	-213.4	-207.7	-202.7	-187.3	-188.4	-197.6	-199.3	-199.6
dev.		-8.2	-2.5	2.5	17.9	16.8	7.6	5.9	5.6
rmsd		12.4	6.1	3.5	10.7	9.4	7.5	5.5	5.7

a Ac=acetyl protecting group,

b The column SCC-DFTB indicates the property of the entire molecule calculated with SCC-DFTB. For the remaining QM/MM calculations the side chain including the link atom is treated with SCC-DFTB while the backbone is treated using MM. Deviations are relative to the entire molecule computed using SCC-DFTB.

c QM/MM frontier between C^{β} and C^{γ}.

d For the doubly protonated His a hydrogen bridge is formed between the proton attached to N^{δ} and the carbonyl oxygen. Fixing the backbone of the amino acid did not prevent this for all schemes. The deviations hence also reflect the imperfect description of a QM/MM hydrogen bridge.



a b

Figure 4.5: Structure and label of atoms of nucleotides studied here: (a) 2′-deoxythymidine (b) 2′-deoxyadenosine. R=2′-deoxyribose

DNA Bases

To further explore the performance of various link atom schemes with different kinds of QM/MM interfaces, the acidity and basicity of 2′-deoxyadenosine (dA) and 2′-deoxythymidine (dT) are examined, which are representative of DNA bases (Fig. 4.5) with the phosphate groups omitted. In addition, the relative energetics of different tautomers is examined. Deprotonation refers to the removal of one proton at ^{3}N and ^{6}N in dT and dA, respectively, and protonation involves ^{4}O and ^{1}N, respectively. The tautomerization involves a proton exchange between the deprotonation and the protonation sites (e.g., between ^{3}N and ^{4}O for dT). The QM/MM partition occurs at the sugar-base bond, and the base was treated with QM (SCC-DFTB). To illustrate the effect of the sugar ring, QM results for the minimal models which are terminated with a hydrogen atom at R were also included (Fig. 4.5).

Overall, the SLA scheme gives the largest error, as found above for other systems. The trends in the performance of the other link atom schemes, however, are somewhat different from those observed for amino acids, which highlights the importance of the charge distribution at the QM/MM interface. For the deprotonation energies (Tab. 4.9) and proton affinities (Tab. 4.10), EXGR, DLA, and DG give satisfactory results, while the DIV and CHSH schemes, which works well for amino acids (section 4.3.1), give much larger errors on the order of 10 kcal/mol in several cases. For methods using Gaussian blurring, no significant dependence on the blurring width is found.

The tautomerization energy (Tab. 4.11), as expected, depends much less on the link atom scheme presumably due to error cancellation. Here, the energetics are in excellent agreement with the full SCC-DFTB calculation with errors less than 2 kcal/mol, even with SLA. The only exception is for dA with the EXGR scheme, which gives a large deviation of 4.1 kcal/mol, which is even larger than that of the minimal model.

4.3.2 Results for Short-Range Proton Transfer Reactions

Results from the last subsection indicate that large errors in protonation and deprotonation energies may arise if QM/MM interactions at the interface are not handled properly. The encouraging aspect, however, is that properties that conserve the to-

Table 4.9: Deprotonation energies for DNA bases: 2′-deoxyadenine (dA) and 2′-deoxythymidine (in kcal/mol)[a].

	SCC-DFTB	SCC-DFTB (R=H)[b]	SLA	EXGR	DIV	DLA/1.0 Å	DLA/1.7 Å	DG/1.0Å	DG/1.7Å	CHSH
dA	363.4	369.8	382.1	367.4	374.1	365.0	366.2	367.2	367.9	374.6
dev.			-18.7	-4.0	-10.7	-1.6	-2.8	-3.8	-4.5	11.2
dT	351.4	356.2	371.0	354.4	359.4	349.2	351.1	352.5	353.7	355.5
dev.			-19.6	-3.0	-8.0	2.2	0.3	-1.1	-2.3	4.1

Table 4.10: Proton affinities for DNA bases: 2′-deoxyadenine (dA) and 2′-deoxythymidine (in kcal/mol)[a].

	SCC-DFTB	SCC-DFTB (R=H)[b]	SLA	EXGR	DIV	DLA/1.0 Å	DLA/1.7 Å	DG/1.0Å	DG/1.7Å	CHSH
dA	237.2	235.5	256.3	241.4	244.5	236.9	238.3	239.9	241.0	238.0
dev.			19.1	4.2	7.3	-0.3	1.1	2.7	3.8	0.8
dT	233.0	222.7	246.8	228.2	235.3	225.8	227.7	229.0	230.2	231.5
dev.			13.8	-4.8	2.3	-7.2	-5.3	-4.0	-2.8	-1.5

Table 4.11: Energy difference for the tautomers for DNA bases (in kcal/mol): 2′-deoxyadenine (dA) and 2′-deoxythymidine[a].

	SCC-DFTB	SCC-DFTB (R=H)[b]	SLA	EXGR	DIV	DLA/1.0 Å	DLA/1.7 Å	DG/1.0Å	DG/1.7Å	CHSH
dA	12.0	11.7	11.5	7.9	11.7	12.9	12.4	12.2	12.2	12.0
dev.			0.5	4.1	0.3	0.9	0.4	0.2	0.2	0.0
dT	5.1	5.9	6.7	6.4	6.1	5.5	5.5	5.5	5.7	5.9
dev.			1.6	1.3	1.0	0.4	0.4	0.4	0.6	0.8

a The column SCC-DFTB indicates the property of the entire molecule calculated with SCC-DFTB. For the remaining QM/MM calculations the QM zone encompasses the purine or pyrimidine ring and the link atom, the sugar is treated using MM. Deviations are relative to the full molecule computed using SCC-DFTB.
b Minimal model.

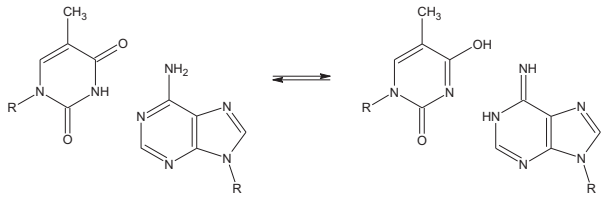


Figure 4.6: Tautomeric proton transfer in a 2′-deoxyadenine (dA) - 2′-deoxythimidine (dT) Watson-Crick base pair.

tal charge (e.g., rotational energy profile and tautomerization energy of DNA bases) show much less dependence on the link atom schemes due presumably to error cancellations. It is interesting to examine the effect of QM/MM frontier treatment on *reaction energetics*, which are the quantity of primary interest in most QM/MM applications. For this purpose, we first examine results for proton transfer in a DNA base pair in the gas phase and then move on to more realistic enzyme systems.

Proton Transfer in a DNA base pair

Proton transfer (PT) was studied for an 2′-deoxyadenine (dA) – 2′-deoxythimidine (dT) Watson-Crick base pair as shown in Fig. 4.6. To simplify geometry optimizations, heavy atoms in the sugar moieties were kept fixed during the PT. The QM/MM partition occurs at the sugar-base bond, and the base was treated with QM (SCC-DFTB). To illustrate the effect of the sugar atoms in PT, full QM results for a minimal model without the deoxyribose residue (R=H) are also included (Fig. 4.6).

The PT is endothermic by about 8 kcal/mol, and has a barrier about 16 kcal/mol without the zero point energy correction (Table 4.12). Similar to the tautomerization energies, the PT energies (both activation barrier and endothermicity) show very minor dependence on the link atom schemes. Indeed, even the minimal models without the sugar part at all give very similar results; the differences for the barrier height ($\Delta E^{\ddagger}$) and endothermicity are 0.5 and 0.9 kcal/mol, respectively. In the QM/MM calculations, the maximum absolute deviation is found for the DIV scheme, which overestimates $\Delta E^{\ddagger}$ by 3.1 kcal/mol and the endothermicity by 2.2 kcal/mol. The absolute deviations of other methods are smaller in magnitude, and all schemes overestimate the energy difference while only the DLA and SLA schemes underestimate the barrier. For example, the DLA scheme underestimates the barrier by 1.1 kcal/mol.

Since the sugar ring does influence the deprotonation and protonation energetics rather significantly (Tables 4.9,4.10), the minimal dependence of the PT energetics on the QM/MM link atom scheme is interpreted as a simple consequence of error cancellation, similar to the case of tautomerization discussed above. According to the Hammond postulate, it is not surprising to see that errors in the barrier height also follows the trend in the reaction energy; this is very reassuring, however, for the application of QM/MM schemes to chemical reaction problems.

Table 4.12: Activation energies $\Delta E^{\ddagger}$ and endothermicities ΔE (in kcal/mol) for the tautomeric proton transfer in the dT-dA base pair[a].

	$\Delta E^{\ddagger}$	ΔE
SCC-DFTB	15.8	8.0
SCC-DFTB (R=H)	14.9	8.5
SLA	15.2	8.2
EXGR	16.6	9.1
DIV	18.9	10.2
DLA/1.0Å	15.0	9.0
DLA/1.7Å	14.7	8.4
DG/1.0Å	17.3	10.2
DG/1.7Å	17.4	10.0
CHSH	17.5	9.9

a The column SCC-DFTB indicates the property of the entire system calculated with SCC-DFTB. For the remaining QM/MM calculations the QM zone encompasses the purine and pyrimidine ring and the link atoms; the sugars are treated using MM.
c Minimal model.

Proton Transfer in Enzymes: TIM and MGS

To illustrate the effect of link atom schemes for reactions in enzymes, the first PT step in Triosephosphate Isomerase (TIM) and Methyl Glyoxal Synthase (MGS) was studied. These systems were chosen based on several considerations: Both systems have been well studied by previous experimental [175–184] and theoretical work [28, 105, 172, 173, 185–190] which provided firm grounds for benchmarking different link atom schemes. Although evolved independently, TIM and MGS have very similar active sites that bind to the same substrate, dihydroxyacetone phosphate (DHAP) and are expected to undergo the same first PT step [191, 192] (Fig. 4.7). The difference is that the general acid in this step is Glu165 for TIM and the shorter Asp71 for MGS; the electrostatic environment is also quite different in the two active sites, which was proposed to be crucial for their functional specificity [64, 105]. Therefore, comparing the dependence of PT energetics to different QM/MM link atom schemes in these enzymes will clearly reflect the range of variations of QM/MM results in realistic biological systems.

As described in Sec. 4.2.3, the QM region, treated with SCC-DFTB, included the substrate, the general base and distal histidine (His95 in TIM and His98 in MGS). Considering the size of the system, instead of performing QM/MM calculations with a much larger QM region, we evaluated the performance of QM/MM frontier treatments by focusing on variations between results obtained from different link atom schemes.

As discussed in previous work, the first proton transfer in both TIM and MGS, despite being catalyzed by the enzymes, is endothermic by a substantial amount [28, 105, 172, 173]. For TIM, the barrier heights range from 10.1 kcal/mol for DIV and 14.3 kcal/mol for DLA/1.7 Å (Table 4.13). These schemes also give the minimum (8.8 kcal/mol) and maximum (13.5 kcal/mol) for the energy difference between the

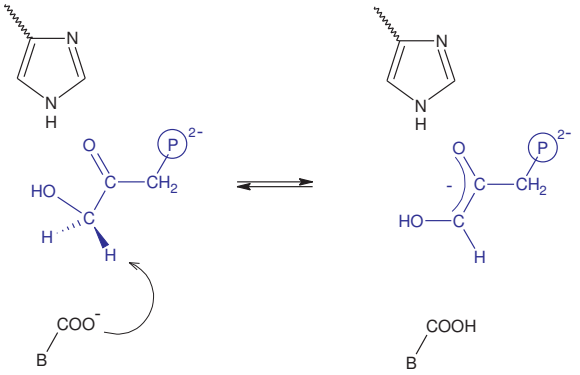


Figure 4.7: First proton transfer step in TIM and MGS, in which the base B abstracting the proton from DHAP is Glu165 and Asp71 for TIM and MGS, respectively; the histidine residue is His95 and His98 in TIM and MGS, respectively. Ⓟ symbolizes the phosphoester group $O\text{-}PO_3^{2-}$

Table 4.13: Activation energies $\Delta E^{\ddagger}$ and endothermicities ΔE (in kcal/mol) for the first proton transfer step in TIM.

	$\Delta E^{\ddagger}$	ΔE
SLA	10.9	8.7
EXGR	12.3	11.0
DIV	10.1	8.8
DLA/1.0Å	14.2	13.4
DLA/1.7Å	14.3	13.5
DG/1.0Å	12.0	10.8
DG/1.7Å	11.6	10.4
CHSH	12.3	11.1

Table 4.14: Activation energies $\Delta E^{\ddagger}$ and endothermicities ΔE (in kcal/mol) for the first proton transfer step in MGS.

	$\Delta E^{\ddagger}$	ΔE
SLA	9.8	7.4
EXGR	8.8	7.0
DIV	11.7	10.2
DLA/1.0Å	12.7	12.2
DLA/1.7Å	12.8	11.9
DG/1.0Å	11.8	10.5
DG/1.7Å	11.9	10.8
CHSH	11.6	10.3

reactant and product. All other methods give energies within these ranges. In all cases the barrier is only slightly higher than the product.

These results are in qualitative agreement with previous work. For example, Cui et al. found a barrier of 17 kcal/mol and an energy difference of 9 kcal/mol using a specifically adjusted AM1 method [28]. More recently, using HF/3-21G as the QM method, the pseudo-bond frontier treatment [150] and more extensive reaction path searches, Liu et al. found the barrier and endothermicity to be 21.3 and 19.4 kcal/mol, respectively [193]. Results from an approximate QM/MM free energy perturbation approach gave very similar results as well [194].

As to critical geometrical parameters in the reactant, transition state and product of the PT, all link atom schemes give very similar results (see Fig. 4.8). For instance, the distances between the transferring proton and the donor/acceptor in the transition state are within 0.05 Å among all frontier treatments. Therefore, the trend that all

link atom schemes yield comparable barriers and energy differences is also reflected in the geometries.

In MGS (Table 4.14), which involves a shorter general base for the PT (Asp71), EXGR and DLA/1.7 Å schemes give the minimum and maximum for the barrier height (8.8 kcal/mol, 12.8 kcal/mol); for the energy difference, EXGR and DLA/1.0 Å give the minimum and maximum values (7.0 kcal/mol, 12.2 kcal/mol). In other words, the range of deviation in the QM/MM results for the barrier heights is approximately the same in MGS (4.0 kcal/mol) and in TIM (4.2 kcal/mol). Critical geometrical parameters are also very similar among all link atom schemes (see Fig. 4.9) Apparently, the shorter QM general base and more polar active site in MGS does not strongly increase the error bars for different link atom approaches. We note that in contrast to the PT in the DNA base pair discussed above, only one of the QM groups is directly influenced by the QM/MM interface; i.e., the entire substrate is treated with QM, and the QM/MM partition explicitly involves only the general base. Therefore, the degree of error cancellation for the PT between the substrate and the general base is expected to be different from that in DNA bases, and it is satisfying to see that relatively small differences between different QM/MM link atom schemes occur for both systems.

4.4 Concluding Discussions

Although numerous studies have demonstrated that QM/MM methods are powerful tools for studying reaction mechanisms [27, 31, 62, 105] and related properties [195–198] in complex systems, many technical details such as the treatment of long-range electrostatics [121, 122, 199], QM/MM van der Waals interactions [108, 200] and extent of configuration samplings [104, 125] as well as the choice of MM partial charges [109, 158] *may* have a major impact on whether the results are meaningful. Precisely how important a contribution might be, which depends on the system under study and questions being addressed, needs to be established by careful and relevant benchmark calculations. Here, we systematically evaluated the importance of the treatment of QM/MM interface (frontier), an issue that has been raised repeatedly in the QM/MM community. As emphasized in a recent study of QM/MM van der Waals interactions [108], to evaluate whether this is the case in *typical* applications, one needs to investigate not only gas phase systems as often done but also realistic condensed phase systems and processes (e.g., reaction in enzymes). Accordingly, we systematically studied the effect of different link atom schemes in QM/MM calculations of proton affinity, deprotonation energy and rotational barriers in a number of gas phase systems of various polarity (alcohols, carbonic acids, amino acids and model DNA bases) as well as proton transfer reactions in model DNA base-pairs and two enzymes. The breadth of the test cases was chosen to ensure a critical evaluation on the performance of QM/MM link atom schemes for realistic systems of biological interest. To focus on the effect of QM/MM frontier treatment, an approximate density functional theory (SCC-DFTB) was used throughout; in most cases, full SCC-DFTB results were taken as reference values for QM/MM results. However, we expect that the trends observed here concerning the effect of QM/MM frontier treatment apply to other QM/MM methods, which is supported

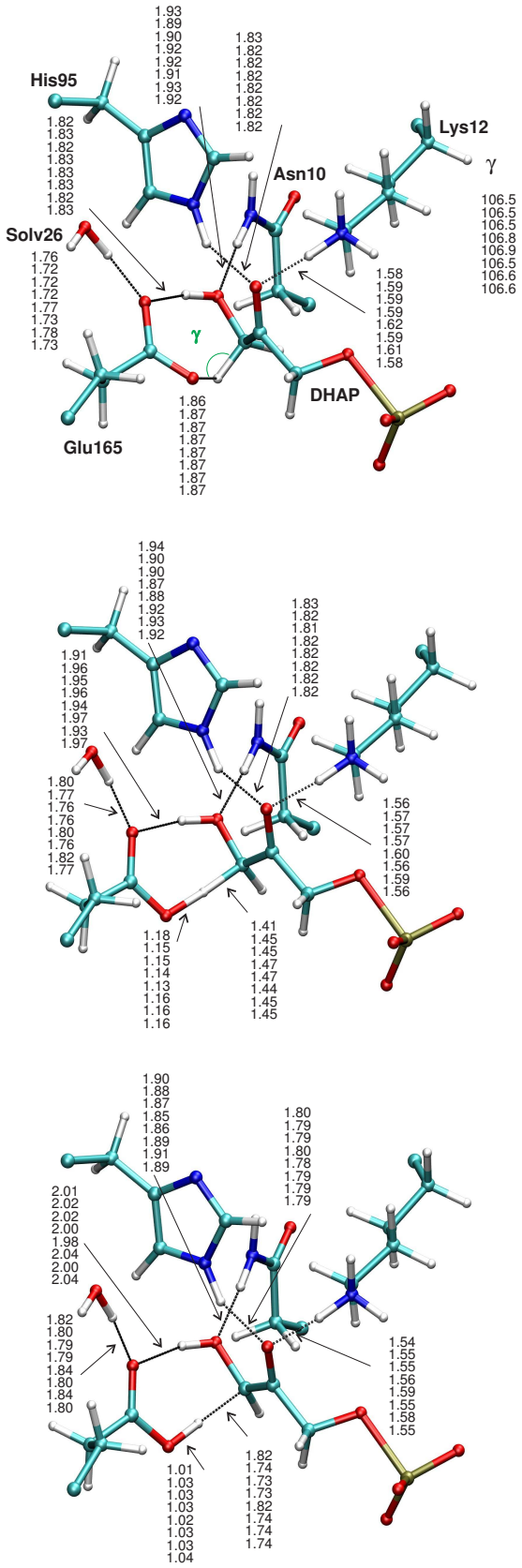


Figure 4.8: Structures and important geometrical parameters for the reactant, transition state and product for the first proton transfer in TIM. Distances are in Å, angles in °. Numbers from *top to bottom*, for SLA, EXGR, DIV, DLA/1.0Å, DLA/1.7Å, DG/1.0Å, DG/1.7Å and CHSH.

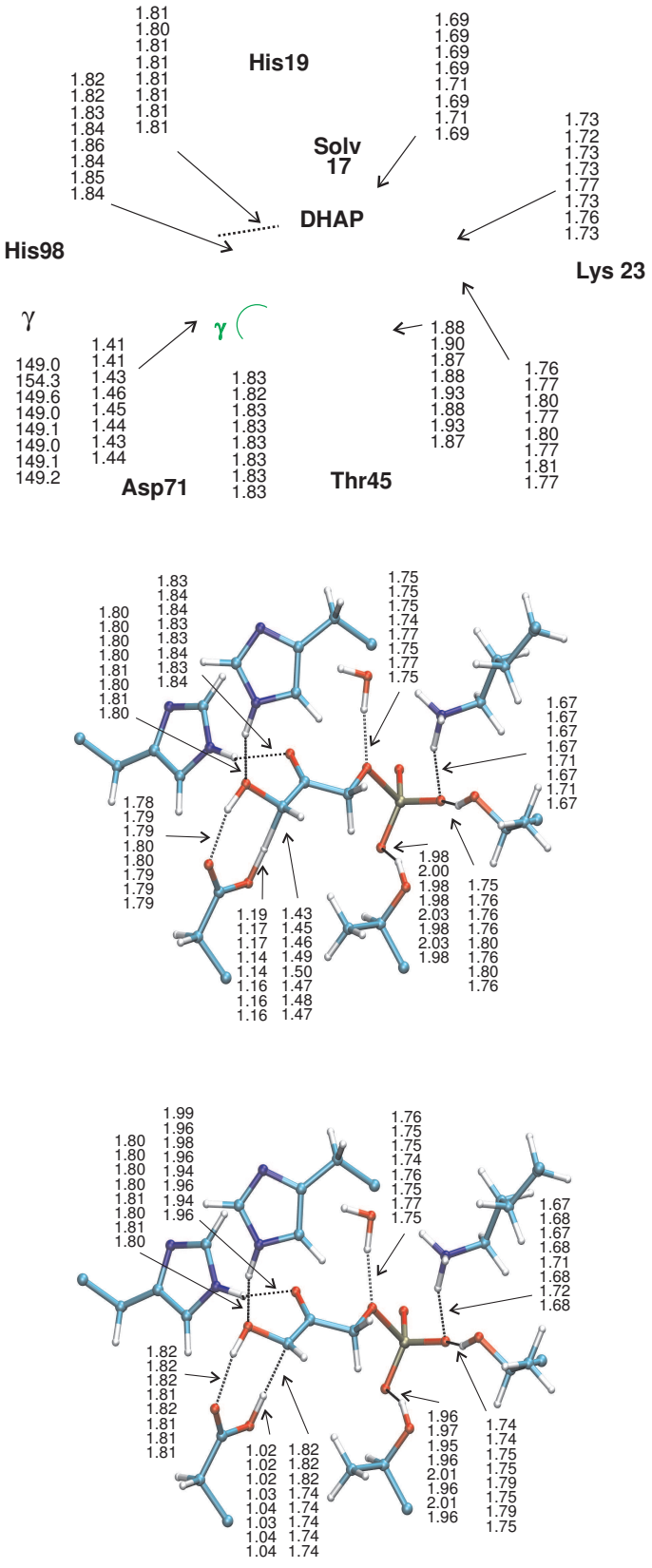


Figure 4.9: Structures and important geometrical parameters for the reactant, transition state and product for the first proton transfer in MGS. Distances are in Å, angles in °. Numbers from *top to bottom*, for SLA, EXGR, DIV, DLA/1.0 Å, DLA/1.7 Å, DG/1.0 Å, DG/1.7 Å and CHSH.

by comparisons to previous studies [115, 151, 160] for several gas phase systems with *ab initio* or DFT as the QM method.

For deprotonation energies and proton affinities, here, we found similar results as previous studies [115] in that the commonly used SLA approach, which (unfortunately) is the default option in CHARMM, gives large errors typically in the range of 15 to 20 kcal/mol [115]. More alarmingly, this error does not decrease systematically as the QM/MM interface moves away from the protonation/deprotonation site (i.e., with increasing QM size). Therefore, for absolute proton affinity and deprotonation energy calculations, such as pK_a evaluations [69], the SLA approach should be avoided. We expect similar trends for other processes that involve significant changes in charge distribution, such as redox potential calculations [65, 68]. In other QM/MM frontier treatments, by contrast, moving the boundary away from the reactive site generally improves proton affinities and deprotonation energies, as also pointed out by other authors [109, 158, 160].

A major problem with SLA is that with the MM host atom (MMHA) excluded from the QM region, a net partial charge is created in the vicinity of the QM region, which causes spurious electrostatic interactions. This problem seems to be alleviated by most other schemes, which conserve the local charge neutrality at the QM/MM interface by either shifting the MMHA charge to the entire host group (CHSH, DIV) or excluding the entire host group (EXGR) Alternatively, blurring MM point charges into Gaussian of finite width as in DG and DLA also works well. It is difficult to argue which among these schemes works consistently the best, since they all have successes and failures among the test cases examined here, which highlighted the importance of including benchmark systems of diverse chemical nature. For example, the EXGR approach works well when the MM host group does not bear any significant dipole moment but may cause serious error otherwise. A highly relevant case involves partitioning amino acids (or proteins) across the C^{α}-C^{β} bond, where EXGR may cause significant error because a major dipole associated with the backbone N-H would be excluded from QM/MM electrostatic interactions (Sec. 4.3.1). For amino acids with sufficiently long side chains (e.g., lysine and glutamate), it seems better to partition across C^{β}-C^{γ}, as suggested in another study [160]. The CHSH and DIV schemes (note that with aliphatic MM host groups, DIV is equivalent to EXGR due to the group definition in the CHARMM force field) gave encouraging results for amino acids, but substantial errors on the order of 10 kcal/mol in deprotonation energy were also observed in model DNA bases. On average, DIV and CHSH give similar results, although the DIV scheme tends to give better deprotonation energies. Methods that rely on Gaussian blurring (DLA and DG) generally give better results than SLA, but the accuracy depends quite sensitively on the blurring width although the dependence decreases as the QM/MM interface moves away from the reactive site (Table 4.4). For dipole moments, significant error may arise with small blurring widths (Table 4.6). With systematic tests on the blurring width, the DLA approach appears to be a promising choice for practical calculations of properties such as proton affinity and deprotonation energy.

The encouraging finding here, although somewhat anticipated, is that reaction energies have a much less pronounced dependence on the link atom scheme. For the tautomerization energy of model DNA bases, proton transfers in model DNA bases pairs as well as in two enzymes, *all* link atom schemes gave very similar results, in-

cluding the SLA approach. Apparently error cancellation is a major reason behind the reduced sensitivity towards QM/MM frontier treatment, because proton affinity and deprotonation energy depend sensitively on the link atom schemes. We note that although this is encouraging in general, there are still non-negligible variations in the results of various link atom schemes. For proton transfers in MGS and TIM, for example, the range is on the order of 2-4 kcal/mol, and whether this is acceptable depends on the specific application.

Overall, we can conclude that while neither of the approaches examined here is a perfect solution to the frontier problems in QM/MM simulations, for many cases both DIV and the DLA scheme give reasonable results, especially for chemical reactions. Therefore, we agree with a previous study [106] that the effect of the frontier treatment on QM/MM study of chemical reactions may be rather small, *if* the total charge is conserved during the reaction.

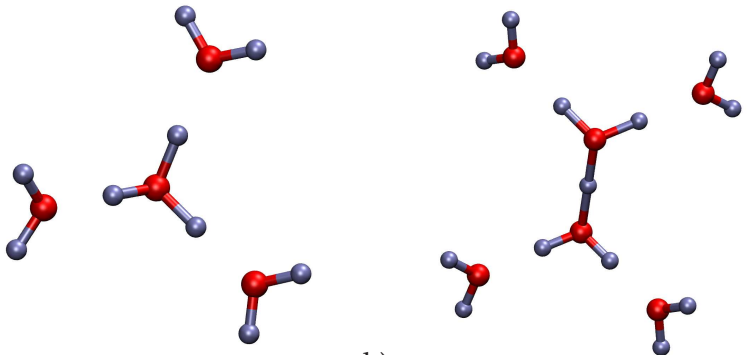
Chapter 5

A new reaction coordinate for proton transfer

5.1 Introduction

Although in some systems PT is fast and thus can be monitored with unbiased nanosecond MD simulations [201], most PT processes have significant barriers and require alternative methods for investigation. Powerful simulation methods such as the transition path sampling technique [202] have been proposed to study the real-time dynamics of rare events in the condensed phase. However their high computational cost limits their applicability. An alternative for studying rare events, e.g., PT, is to compute the PMF along a well-chosen reaction coordinate. By "well-chosen", we mean that the reaction coordinate captures the nature of the degree(s) of freedom that most strongly regulate the reaction kinetics. Due to charge migration in long-range PT, the protein and solvent environment undergo significant reorganizations; an appropriate reaction coordinate, therefore, is the energy gap between different diabatic states that correspond to localized proton coordinations, similar to the "solvent coordinate" used to describe electron transfers in solution. This energy gap coordinate was indeed used extensively by Warshel and co-workers [26, 41, 125] in the study of long-range PT using EVB potentials [103], in which the diabatic states are well defined. With adiabatic QM/MM potentials, however, it is more difficult to define an energy gap coordinate; moreover, application of the energy gap coordinate to simultaneous (concerted) multiple proton transfers is less straightforward.

Therefore, we choose to define the reaction coordinate in terms of the geometrical and charge property of the PT reaction, i.e., the location of the center of excess charge (CEC) relative to the proton donor and acceptor groups. We note that although this type of coordinate is geometrical in nature, its dependence on the charge distribution of the "proton wire" ensures that the environment appropriately reorganizes as its value varies. Indeed, previous comparison of the energy gap coordinate and a geometrical coordinate found very similar PMF results for short-range proton transfer reactions. [203]. Moreover, activated dynamics simulations obtained a transmission coefficient fairly close to unity (~ 0.4) for a number of studies of short-range PT reactions in enzyme systems using geometrical coordinate to define the reaction



a) b)

Figure 5.1: Idealized structures of the a) Eigen- and b) solvated Zundel-ion

coordinate[204, 205]. Whether the success applies quantitatively to long-range PT processes remains to be investigated.

Generally speaking, when defining a geometric reaction coordinate, the task is to extract those degrees of freedom best describing the reaction. Specifically for PT involving water molecules, a rapid interchange takes place [206] between structures which are close to the ideal structures of Eigen- and Zundel ions [207, 208] (Fig. 5.1). The defect, i.e. the excess proton coordinated to water molecules, travels without any atom actually moving further than fractions of an Ångstrøm [20, 206]. Hence the challenge for long-range PT is to separate these subtle fluctuations responsible for proton transfer from other motion in the immediate protein and solvent environment.

5.2 Previous suggestions

anti-symmetric stretch coordinate

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),

$$\delta = r^{D,H} - r^{A,H} \tag{5.1}$$

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[42], its use becomes cumbersome and less robust for more complicated pathways.

center of excess charge (CEC)

In the study of PT through the water wire in gramicidin A, a reaction coordinate based on the *center of excess charge (CEC)* was used by Roux and coworkers [45, 209], which involves the projection of the total dipole moment of the water wire on the

z-axis (the axis of the water wire),

$$\mu_z = q_H \sum_{i=1}^{N_H} r_z^{H_i} + q_O \sum_{j=1}^{N_O} r_z^{O_j} \tag{5.2}$$

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. 5.2a) the CEC gives the z component of the total dipole moment; in a protonated wire (e.g. $O_{10}H_{21}^+$, Fig. 5.2b-e), on the other hand, it corresponds to the projection of the center of excess charge (proton defect) 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. [17] (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,

$$\nu = \frac{\sum_{i=1}^{N_O} r_z^{O_i} w^{O_i}}{\sum_{O_i} w^{O_i}} \tag{5.3}$$

$$w^{O_i} = \left[\sum_{j=1}^{N_H} f_{sw}(d_{O_i,H_j}) \right] - 2 \tag{5.4}$$

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[17]:

$$f_{sw}(d) = \frac{1}{1 + exp\left[(d - r_{sw})/d_{sw}\right]} \tag{5.5}$$

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 can't 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. 5.5), 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.

Table 5.1: Values for different reaction coordinates (in Å) for the water wires displayed in Fig. 5.2.

	μ_z/e	ν	ξ_z
	Eq. 5.2	Eq. 5.3	Eq. 5.6
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	-4.5	-4.2

5.3 A new reaction coordinate

modified center of excess charge (mCEC)

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

$$\xi_z = \sum_{i=1}^{N_H} r_z^{H_i} - \sum_{j=1}^{N_X} w^{X_j} r_z^{X_j} - \sum_{i=1}^{N_H}\sum_{j=1}^{N_X} f_{sw}(d_{X_j,H_i}) \cdot \left(r_z^{H_i} - r_z^{X_j}\right) \tag{5.6}$$

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. 5.2, the oxygen atoms have a w^X of +2 (also see Sect. 3.1 for a more general example). In the more general case, X can be any proton acceptor with a different weight (see below for more specific examples). The correction (third term in Eq. 5.6) consists of the sum over all contributions to the z-component of the bond vector from individual bonds. To establish the definition of a bond in this context, we use the same switching function $f_{sw}(d)$ as Chakrabarti et al. (Eq. 5.5) for the distance d_{X_i,H_j}. Parameters of r_{sw} =1.15 Å and d_{sw} =0.045 Å were found to give the best results.

5.4 Comparison of different coordinates for linear proton wires

To illustrate the difference between the newly proposed ξ_z coordinate and previous suggestions (μ_z/e, ν), we compare their behavior with different situations of protonated and non-protonated water wires (Fig. 5.2). The coordinates were collected from molecular dynamics simulations of the model channels described below (Sect.3.2). When there is no excess proton in the water wire, the value of the CEC

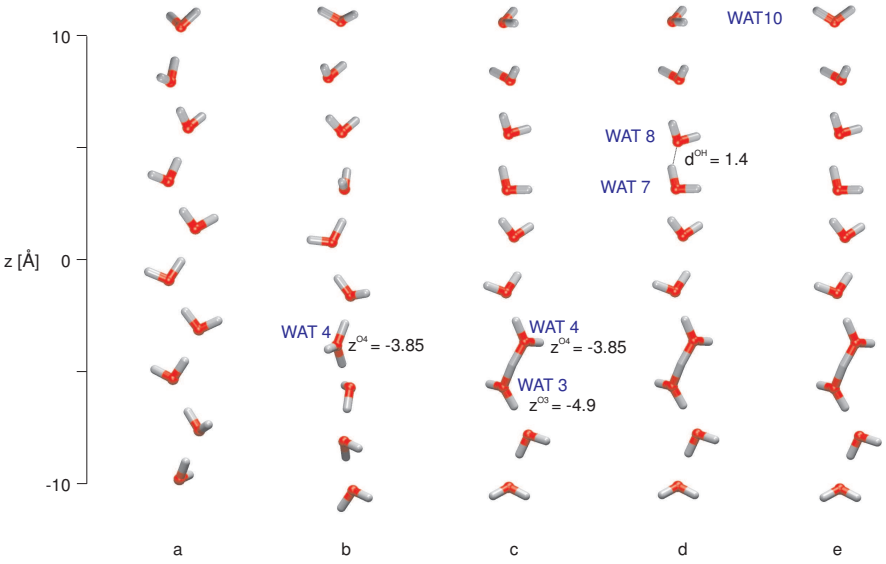


Figure 5.2: Water wire models for illustrating different reaction coordinates that describe long-range proton transfers. a) a water wire without any excess proton; b) a water wire with a H_3O^+-ion located at WAT4; c) a water wire with a Zundel $H_5O_2^+$-ion between WAT3 and WAT4; d) same as c) but 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 Tab. 5.1. The distance cutoff for the bonds drawn was chosen to be 1.3 Å.

coordinate (μ_z/e) is significantly different from 0 as expected (Tab. 5.1). The value of the mCEC coordinate (ξ_z), by contrast, is very close to zero; this demonstrates that the correction (third term in Eq. 5.6) appropriately removes contributions from any components irrelevant to the proton transfer as designed. The reaction coordinate of Chakrabarti et al. [17] (ν) gives unstable results (large numerical value) for unprotonated water wire because the denominator in Eq. 5.3 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. 5.2b), 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 towards WAT3 (Table 5.1). With a Zundel like ion in the wire (Fig. 5.2c), 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. 5.2d, 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 coordinate changes significantly: the oxygen in WAT8 appears to have more than two bonded hydrogens, leading to a non-vanishing 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. 5.2. 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. 5.2e). 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. 5.2b) to a Zundel ion (Fig. 5.2b). This gets even more problematic when larger systems with more water molecules are examined.

5.5 Generalization to complex proton wires in three dimensions

Test calculations in the last subsection demonstrated that the mCEC coordinate, ξ_z, is robust for describing proton transfer in a linear water chain. Although this is sufficient for long-range PT in some membrane channels with nearly ideal transfer

geometry, it is useful to extend ξ_z to a three-dimensional vector such that PT along more complex pathways can be treated. The proposed functional form is,

$$\boldsymbol{\xi} = \sum_{i=1}^{N_H} \boldsymbol{r}^{H_i} - \sum_{j=1}^{N_X} w^{X_j} \boldsymbol{r}^{X_j} - \sum_{i=1}^{N_H} \sum_{j=1}^{N_X} f_{sw}(d_{X_j,H_i}) \cdot \left(\boldsymbol{r}^{H_i} - \boldsymbol{r}^{X_j} \right) \tag{5.7}$$

where notations are similar to those in Eq. 5.6 except for the trivial vector extension.

The discussions thus far assumed that each group along the PT pathway can either accept or donate a proton, which is reasonable for water molecules. For long-range PT in proteins, titratable groups may participate in the PT pathway, which may give rise to more complex scenarios that further complicate the definition of the reaction coordinate. For example, in the sidechains of glutamate and aspartate, the protonation and deprotonation of the two titratable heavy atoms X_k,X_l (e.g., Oϵ atoms in Glu) are coupled. In cases, where the final donor or acceptor is a carboxylic, it can be desirable to treat the two carboxylic oxygens equivalently [1].

In those cases, a further revision of the mCEC is proposed: the weights for the donor/acceptor pair in (Eq. 5.7) are set as: $w^{X_k} = w^{X_l} = w_{pair}^{X_k,X_l}/2$, where w_{pair} is the number of protons coordinated to the *residue* in its reference state (as described above for atoms). Moreover, the following term is added to $\boldsymbol{\xi}$ for each donor-acceptor pair:

$$\boldsymbol{\xi}'_{pair}(X_k, X_l) = \frac{1}{2} \left[m(X_k, \{H\}) \cdot \left(\boldsymbol{r}^{X_l} - \boldsymbol{r}^{X_k} \right) + m(X_l, \{H\}) \cdot \left(\boldsymbol{r}^{X_k} - \boldsymbol{r}^{X_l} \right) \right] \tag{5.8}$$

The term $m(X, \{H\})$ contains the information regarding whether there is at least one proton coordinated to the respective atom. It is a differentiable approximation to the maximum function for the switching functions ($f_{sw}(d_{X,H_i})$) concerning distances between atom X and all protons along the PT pathway:

$$m(X, \{H\}) = \sum_{H_i \in \{H\}} f_{sw}(d_{X,H_i})^{n+1} \Bigg/ \sum_{H_i \in \{H\}} f_{sw}(d_{X,H_i})^n \tag{5.9}$$

where a reasonably large integer (e.g., 15) works well for n.

To illustrate the physical significance of the terms in Eq. 5.8,5.9, consider the sequence of structures depicted in Fig.5.3a)-c). With idealized geometry in structures a,b and c, the weights $m(X_k, \{H\})$ and $m(X_l, \{H\})$ are (0,1), (1,1) and (1,0), respectively. In the transitional structure shown in Fig. 5.3b, the mCEC with the correction term $\boldsymbol{\xi}'_{pair}$ included locates the excess proton at the midpoint between the two oxygen atoms. The switching functions in Eq. 5.8 ensure a smooth transition for the reaction coordinate during the proton transfer. The maximum like function m has to be used to allow an alternative PT mechanism as shown in Fig. 5.3d, which is conceivable in the study of the bacterial photosynthetic reaction center (proton transfer through Asp213 (L) and Ser223 (L) to Q_B). Here, $m(X_k, \{H\})$ and $m(X_l, \{H\})$ are

[1] This case was not considered in the original paper [124]. The prefactor in Eq. 5.8 was hence adjusted.

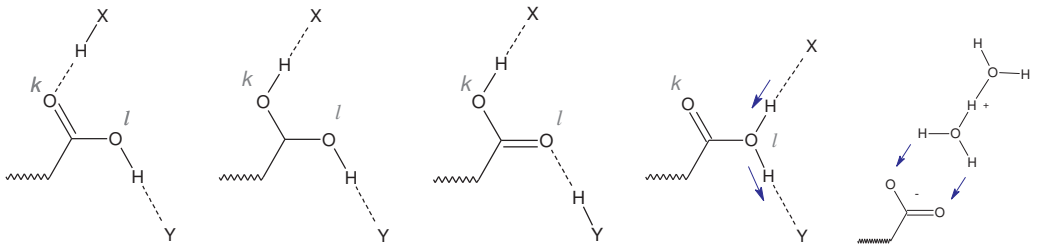


Figure 5.3: (a-d) Proton transfer through a protonated carboxylic acid as an example of a coupled donor-acceptor pair (O_k,O_l). Two mechanisms are shown: (a)-(c) sequence of steps, with O_k and O_l acting as proton acceptor and donor respectively. (d) O_l acts both as donor and acceptor (e) proton transfer to two equivalent acceptors in a carboxylic acid. Arrows indicate the movement of the protons during the reaction.

0 and 1 respectively. In some cases, the structure of the proton wire allows proton transfer to either oxygen atoms of a carboxylic acid as shown in Fig. 5.3e. To treat the protonation at either of the sites as equivalent, the addition of the term in Eq. 5.9 will locate the excess proton at the midpoint of the two carboxylic oxygen atoms.

Unlike ξ_z, the vectorial quantity $\boldsymbol{\xi}$ can not be readily used in PMF calculations. For this purpose, we further define a collective reaction coordinate ζ for mapping the transfer of the mCEC (as reflected by $\boldsymbol{\xi}$) between an initial donor atom D and a final acceptor atom A,

$$\zeta = -\frac{1}{1+exp\left[Cd_{\xi,D}\right]} + \frac{1}{1+exp\left[Cd_{\xi,A}\right]} \tag{5.10}$$

where $d_{\xi,D}$ and $d_{\xi,A}$ represent the distance from mCEC to the initial donor and the final acceptor atoms respectively; C is a coefficient defining the range of the reaction coordinate and is taken as 1 unless otherwise stated. Apparently, ζ is a switching function dominated by $d_{\xi,D}$ and $d_{\xi,A}$ at the lower and higher limits, respectively; as ζ evolves from negative to positive values, the mCEC displaces from the initial donor to the final acceptor. The choice of functional form in Eq. 5.10 is not unique and many other smooth switching functions could be used; e.g., another possible expression is [210],

$$\zeta_R = \frac{d_{\xi,D} - d_{\xi,A}}{d_{\xi,D} + d_{\xi,A}} \tag{5.11}$$

The value of ζ_R of -1 and 1 corresponds to the mCEC localized to the initial donor and final acceptor, respectively.

Finally, we note that due to the collective nature of ζ and ζ_R, their values do not explicitly depend on the number of shuttling groups (i.e., water and titratable residues) along the PT pathway. Therefore, they are conveniently suited for studying PT in complex environments, in which there are either many possible transfer pathways with different numbers of shuttling groups existing at the same time or the number of shuttling groups fluctuates at a time-scale faster compared to that of

the PT process. A good example is carbonic anhydrase, for which water wires of different lengths were consistently observed in different MD simulations[43, 44, 122] and the typical life-time for such water wires is on the order of pico-seconds, as compared to the μs time-scale for the PT.

5.6 Test calculations: Application to proton transfer in carbonic anhydrase II

Here, two PT pathways in a realistic enzyme model for carbonic anhydrase were studied to test the robustness of new reaction coordinates for long-range PT through non-linear water chains.

Carbonic anhydrase II (CAII) is a zinc-enzyme that has been widely used as a prototypical model for studying long-range PT in proteins. Its biological function is to interconvert CO_2 and HCO_3^-, which makes it important in respiration processes [211, 212].

An important step during the catalytic cycle is the transfer of a proton between a zinc-bound water and a histidine residue (His 64) close to the surface of the protein; in the forward direction, the PT generates a zinc-bound hydroxide, which is the catalytic species that reacts with CO_2. The distance between the zinc-bound water and His 64 is about 8 Å in the x-ray structure, which led to the suggestion that PT has to be mediated by water molecules in the active site [213–215]. Although two-water molecules have been observed in the x-ray structures that connect the zinc-bound water and His 64 through hydrogen bonds, simulation studies with different potential functions [43, 44, 122] have shown that the water structure in the active site is rather dynamical and water bridges connecting the zinc site and His 64 typically include from two to four water molecules. The on-going debate [40, 41, 43, 44, 137, 216] concerns whether formation of a specific type of water bridge contributes dominantly to the kinetic bottleneck of the PT and whether the PT proceeds through a concerted or step-wise mechanism [39–42, 217]. In the current work, we only use CAII to illustrate the new reaction coordinates via comparison to minimum energy path results.

Computational Setup

The enzyme model was set-up using the GSBP protocol: it contained a 16 Å microscopic spherical region centered around Zn^{2+} in which all atoms were allowed to move, and a 2 Å layer of Langevin region in which atoms were harmonically constrained. The remaining protein atoms and x-ray water molecules formed the "outer" region were fixed in space. Spherical harmonics up through the 20th order were used as the basis functions for the electrostatics in GSBP. The QM region included the zinc ion and all of its ligands (His 94, 96, 119 plus a water), the His 64 sidechain and the bridging water molecules (2 and 4, see below); they were treated using SCC-DFTB, while the rest atoms were described with the CHARMM 22 force field for proteins [58]. For further details, refer to Ref. [42, 122].

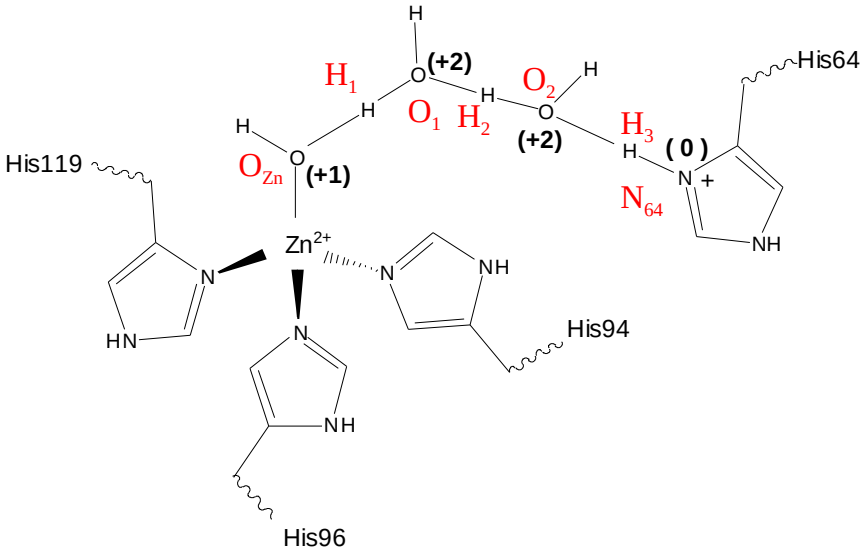


Figure 5.4: Atom labels and weights (in parentheses) associated with the definition of the modified center of excess charge coordinate for CAII computations (Eq. 5.7)

Two configurations collected from MD simulations, as described in an earlier publication [122], were used as the starting geometry. The MD simulations were done in the "COHH" state [43], in which the zinc-bound water is deprotonated and His 64 is doubly protonated ($H64H^+$); thus the PT proceeds from $H64H^+$ to the zinc-bound hydroxide in the present set of results. The two starting configurations contain two and four bridging water molecules, respectively, between the zinc-bound hydroxide and $H64H^+$. Adiabatic mapping calculations were carried out using both ζ (Eq.5.10) and ζ_R (Eq.5.11) to map out the potential energy profile along these approximate reaction coordinates; the weights associated with the heavy atoms along the water wire for ζ and ζ_R are illustrated in Fig. 5.4. Minimum energy pathways were determined using the conjugate peak refinement (CPR) algorithm [55] in CHARMM. The results of the adiabatic mapping and the MEP, both the energetics and critical geometrical parameters along the path, are compared in the following. In addition, adiabatic mapping calculations were carried out for strictly step-wise PTs for the two-water bridge case, to illustrate the importance of having a flexible reaction coordinate. These calculations used a series of anti-symmetric stretch (δ, Eq. 5.1) to drive the sequential PTs; three possible sequences of transfers were studied (see below). No mass-weighting was used in any calculations and no zero-point energy correction was included in the energetics.

Results

For the *two-water bridge configuration* chosen here, the MEP calculations identified only one unique saddle point (Fig. 5.5a) and the barrier height is approximately 8.7 kcal/mol. The transfer of the three protons occurred in a nearly concerted fashion, with the transfer of the middle proton (H_2) somewhat ahead of the other two (Fig. 5.6a, Tab. 5.3). The distances between proton donor and acceptor atoms also

Table 5.2: Comparison of barrier heights $E^{\ddagger}$ and exothermicities ΔE (in kcal/mol) for the proton transfers in carbonic anhydrase obtained using different protocols[a]

	$E^{\ddagger}$	ΔE
2-water bridge		
MEP	8.7	3.0
ζ	9.7	3.2
ζ_R	9.3	3.0
Step-wise (seq. 1)[b]	22.7	5.9
Step-wise (seq. 2)[b]	37.1	5.9
Step-wise (seq. 3)[b]	29.5	7.3
4-water bridge		
MEP	19.6	7.4
ζ	19.0	7.4
ζ_R	19.9	7.4

a. The MEP results are from minimum energy path calculations using the conjugate peak refinement algorithm. Other results were obtained using adiabatic mapping calculations using specific reaction coordinates (ζ (Eq.5.10), ζ_R (Eq.5.11) or δ (Eq. 5.1)). The SCC-DFTB/CHARMM-GSBP protocol was used for the potential function (see text).

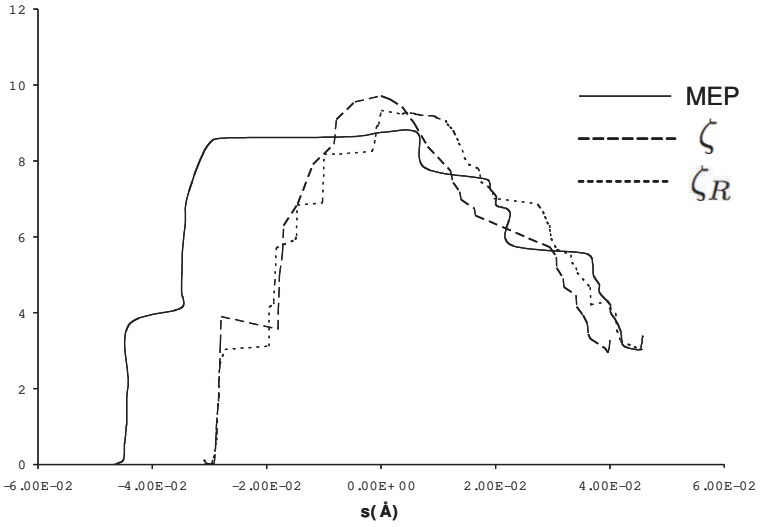
b. In the "Step-wise" calculations, different sequences of PTs were followed in a strict step-wise manner using consecutive adiabatic mapping calculations with δ as the reaction coordinate. Following the notation in Fig. 5.4, the PT sequences are: H_1-H_2-H_3(seq. 1); H_1-H_3-H_2 (seq. 2); H_3-H_2-H_1 (seq. 3). The difference between the exothermicity in step-wise PTs and MEP/ζ,ζ_R results is due to the relaxation in the MM environment after a larger number of minimizations; the small value relative to the difference in barrier heights, however, does not complicate the comparison between different PT pathways.

varied significantly during the PT, which is a well-known result [218] (Fig. 5.6b).

Results from adiabatic mapping using ζ and ζ_R are in excellent agreement with the MEP. As shown in Fig. 5.5a and Tab. 5.2, barrier heights, exothermicity of the reaction and qualitative shape of the energy profile from MEP calculations are well reproduced by both ζ and ζ_R based adiabatic mapping calculations; e.g., the barrier using ζ and ζ_R is 9.7 and 9.3 kcal/mol, respectively, compared to the value of 8.7 kcal/mol from MEP calculations. The evolution of critical geometrical parameters along the path (i.e., those presented in Fig. 5.6) are also well reproduced using both ζ and ζ_R based calculations (Tab. 5.3). This is satisfying considering that ζ and ζ_R are collective coordinates that do not specify, *a priori*, the sequence of proton transfers. Specifically, the structure of the saddle point from MEP calculations agrees well with the highest-energy structures from adiabatic mapping calculations using the two reaction coordinates (Fig. 5.7a). The critical distances shown in Tab. 5.3 further illustrate the concerted nature of the PT transition state.

Adiabatic mapping calculations following a strictly step-wise transfer mechanism, by contrast, showed very different energetics. As shown in Tab. 5.2, the energetics depend very sensitively on the sequence of the transfers. Transfer sequences that leave localized the hydroxide-hydronium pair is highly unfavorable energetically

(a)



(b)

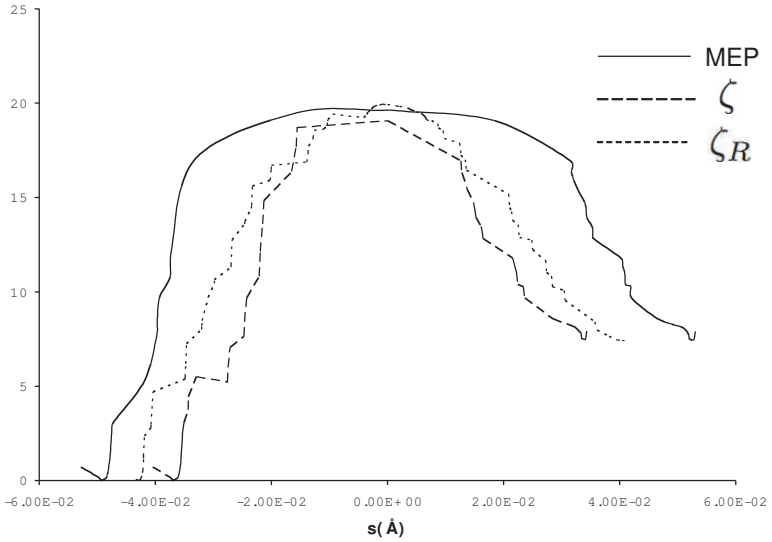
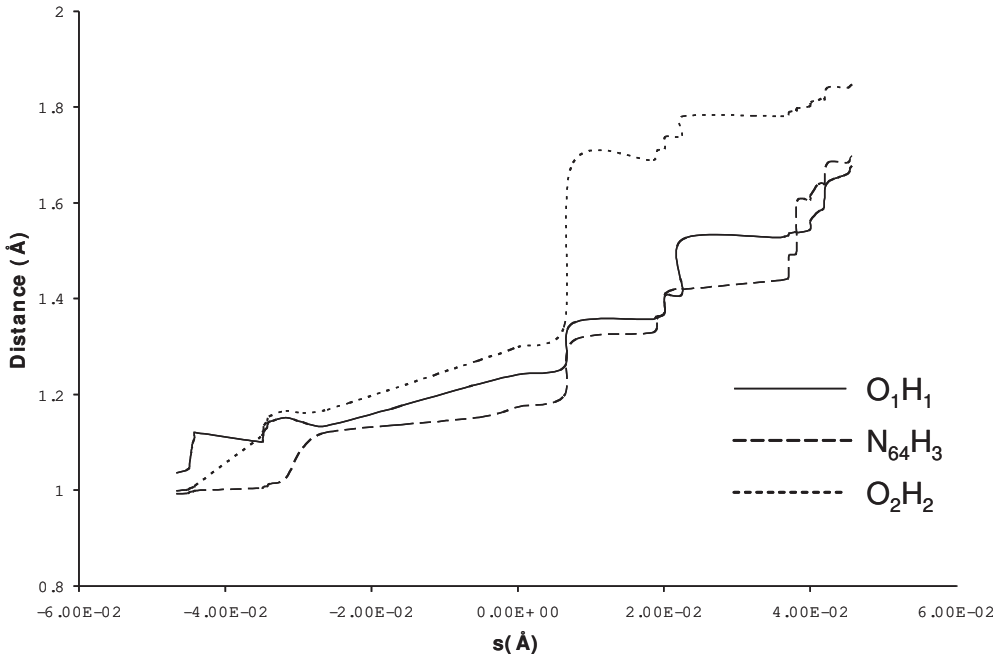


Figure 5.5: Energy profiles along different reaction paths) for the proton transfer in carbonic anhydrase starting from a configuration with (a) two (b) four bridging water molecules. The horizontal axis is defined as the path length (in Å) measured relative to the highest energy structure along the path. The longer path lengths and broad features in the MEP results are due to the small cumulative variations in the MM degrees of freedom, which is a limitation in the conjugate peak refinement algorithm used here.

(a)



(b)

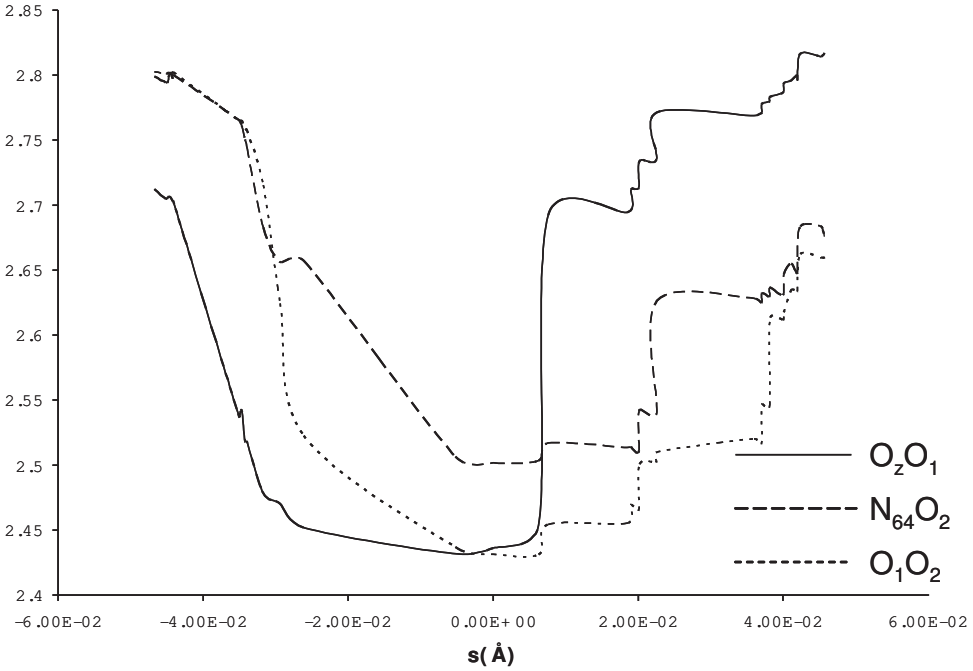


Figure 5.6: Critical geometrical parameters along the minimum energy path for the proton transfer in carbonic anhydrase starting from a configuration with two bridging water molecules. The horizontal axis is defined as the path length (in Å) measured relative to the highest energy structure along the path. (a) The distances between transferring protons and the original oxygen donor atoms; (b) The distances between proton donor and acceptor atoms. For the atom labels, see Fig. 5.4.

with barriers on the order of 40 kcal/mol; a sequence that involves the formation of a localized hydroxide is less demanding although the barrier is still substantially higher (22 kcal/mol) than the concerted pathway ($\sim$ 9 kcal/mol) revealed in both MEP and adiabatic mapping calculations using ζ and ζ_R. It is not the intention here to conclude based on these calculations that concerted PT in CAII is much preferable than a strictly step-wise mechanism, it is worth mentioning that reaction path calculations starting from multiple ($\sim$ 100) initial structures and PMF calculations do support this scenario [42].

For the *four-water bridge configuration* it is striking that only one saddle point was found in the MEP calculations as well. The broad feature of the MEP energy profile in the barrier region is due to the small variations of MM atoms rather than major changes in the reactive degrees of freedom, which is a limitation to the CPR algorithm in the absence of any special coordinate weighting. The barrier height of 19.6 kcal/mol in this specific reaction path is substantially higher than the barrier of 8.7 kcal/mol in the two-water-bridge case discussed above. We emphasize that these results alone do not suggest that the barrier of PT through a longer water bridge is necessarily higher than that through a shorter water bridge, because significant variations in the behavior was found when multiple protein configurations were considered [138]. The issue of interest here is whether ζ and ζ_R can reproduce MEP results when the length of the water wire is long. The results suggest that this is the case, as both energetics (Tab. 5.2) and transition state structures (Fig. 5.7b, Tab. 5.6) were well reproduced by adiabatic mapping calculations; the agreement is more impressive using ζ_R.

In short, test calculations using CAII produced encouraging results suggesting that both collective coordinates introduced here, ζ and ζ_R, can be used to describe the energetics and mechanism of PT spanning a long-distance through water wires of rather complex geometries (Fig. 5.7).

5.7 Concluding Remarks

We introduced a new set of collective coordinates for characterizing the progress of long-range PT. In contrast to earlier suggestions, which were also analyzed here, the new set of coordinates based on the modified center of excess charge (mCEC) works for PT along not only linear but also complex three-dimensional transfer pathways. These coordinates were verified by comparing the corresponding adiabatic mapping results (energetics and key geometrical parameters) to minimum energy path calculations. Excellent agreements were observed for PT through rather complex water wires with realistic model of carbonic anhydrase using a SCC-DFTB/CHARMM potential. These results suggest that the new coordinates can be used as the reaction coordinate for computing meaningful potential of mean force for long-range PT. We note in particular that the collective nature of the new coordinates makes it straightforward to consider the equilibration of different water-wire configurations in the PMF calculations, which is often found rapid compared to the time-scale of the PT process.

Table 5.3: Comparison of critical distances (in Å) in the transition state for the proton transfer through two bridging water molecules in carbonic anhydrase obtained using different protocols[a]

	MEP	ζ	ζ_R
$r_{O_{Zn}O_1}$	2.43	2.44	2.48
$r_{O_1O_2}$	2.44	2.44	2.44
$r_{O_2N_{H64}}$	2.50	2.51	2.58
$r_{O_{Zn}H_1}$	1.26	1.28	1.34
$r_{H_1O_1}$	1.17	1.16	1.14
$r_{O_1H_2}$	1.14	1.16	1.23
$r_{H_2O_2}$	1.30	1.28	1.20
$r_{O_2H_3}$	1.27	1.29	1.43
$r_{H_3N_{H64}}$	1.24	1.23	1.16

a. The transition state for MEP is the saddle point along the minimum energy path obtained using the conjugate peak refinement algorithm. For ζ (Eq.5.10) and ζ_R (Eq.5.11), the structure corresponding to the highest potential energy in the adiabatic mapping results were used. For atom labels, see Fig. 5.4.

Table 5.4: Comparison of critical distances (in Å) in the transition state for the proton transfer through four bridging water molecules in carbonic anhydrase obtained using different protocols[a]

	MEP	ζ	ζ_R
$r_{O_{Zn}O_1}$	2.42	2.45	2.42
$r_{O_1O_2}$	2.44	2.52	2.49
$r_{O_2O_3}$	2.43	2.65	2.43
$r_{O_3O_4}$	2.42	2.57	2.43
$r_{O_2N_{H64}}$	2.49	2.45	2.42
$r_{O_{Zn}H_1}$	1.16	1.15	1.15
$r_{H_1O_1}$	1.26	1.30	1.27
$r_{O_1H_2}$	1.20	1.15	1.20
$r_{H_2O_2}$	1.23	1.31	1.22
$r_{O_2H_3}$	1.17	1.04	1.18
$r_{H_3O_3}$	1.26	1.53	1.25
$r_{O_3H_4}$	1.14	1.03	1.14
$r_{H_4O_4}$	1.30	1.62	1.30
$r_{O_4H_5}$	1.22	1.15	1.22
$r_{H_5N_{H64}}$	1.27	1.37	1.16

a. See footnote of Tab. 5.3. For atom labels, see Fig. 5.4

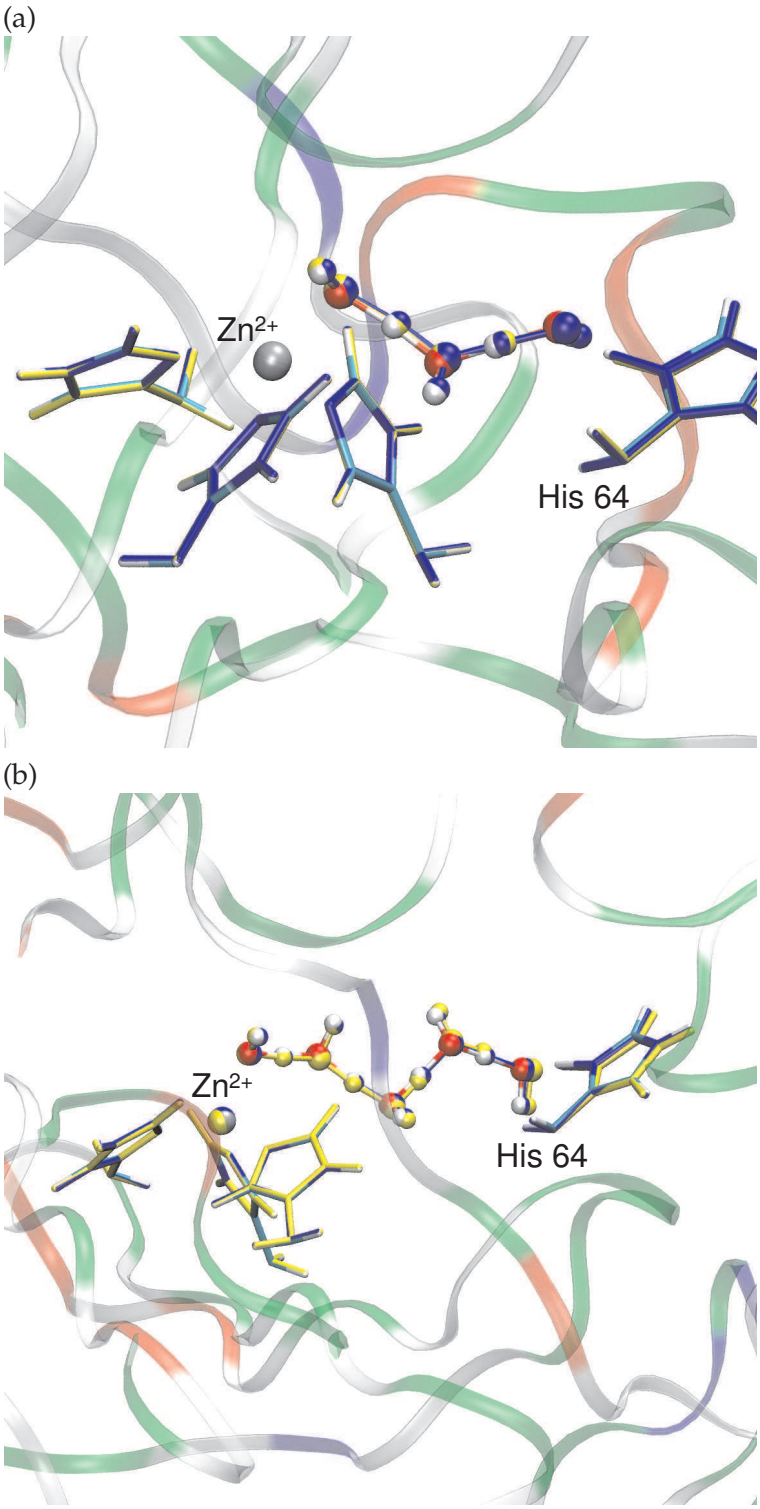


Figure 5.7: Overlay of transition state structures from adiabatic mapping and minimum energy path calculations for proton transfer through (a) two (b) four bridging water molecules in carbonic anhydrase. For precise values for critical geometrical parameters, see Tab. 5.3 and 5.6.

Chapter 6

Proton transfer in the bacterial reaction center

6.1 Introduction

The bacterial reaction center (BRC) is a protein-pigment complex that, through a complex series of events, converts light energies into electrostatic energy in the form of a proton gradient. Although the light can be absorbed directly by the bacterial reaction center *in vitro*, nature has devised a system of light-harvesting proteins (LH1-LH3), that convey the light energy to the bacterial reaction center through excitation transfer, for increasing quantum yield [219][1].

The bacterial reaction center is also a model system which has been used to study photosynthesis as well as more elementary processes in biophysics, such as excitation, electron transfer (ET), and proton transfer (PT). Since the publication of the first structures of the reaction center obtained by x-ray diffraction in *Rhodopseudomonas (Rp) viridis*[2] [6], the research of photosynthesis proteins has received attention by numerous groups. In this work, the treatise will focus on the bacterial reaction center in *Rhodobacter (Rb.) sphaeroides*, as more research data are available for the system in this species.

Structure

The BRC consists of three protein chains labeled L, M, and H, with 281, 307, and 260 residues, respectively. It is embedded in the inner cell membrane of the bacteria, which separates the cytoplasm (contents of the cell) from the periplasm (intermembrane space). The membrane spanning part of the protein consists of 11 α-helices. The lipophylic portion of this transmembrane part was determined to have

[1] For brevity, in the following, it is assumed, that the special pair *D* (vide infra) is excited directly as is the case in most experimental settings.

[2] It was proposed that the species *Rhodopseudomonas viridis* should be repositioned in the phyllogenetic tree and renamed into *Blastochloris viridis* based on genetic comparisons with closely related species [220].

a length of about 35 Å, which is in good agreement with the membrane thickness in *Rb. sphaeroides*, determined to be about 40-45 Å [27].

A number of cofactors are bound non-covalently to the the protein (Fig. 6.1) in two branches (A and B) with near C_2-symmetry: A dimer of two bacteriochlorophylls (special pair D, two secondary bacteriochlorophylls (BCl), two bacteriopheophytins (BPh), a non-heme iron (Fe^{2+}) and two ubiquinones (Q_A and Q_B). The structure of the two ubiquinone molecules is depicted in Fig. 6.2. The role of the ferrous ion is not entirely clear. While the metal ion seems to have a stabilizing effect on the protein structure, the electronic properties do not seem to play an essential role for the functioning of the protein. Other divalent metal ions can substitute it without loss of functionality [221]. Despite the symmetry of the protein and the bound prostethic groups, generally only the A-branch is photoactive. This preferential direction may be due to the different polarities of the respective binding pockets for the quinones [222].

Photocycle

Overall, the events in the photocycle of the bacterial reaction center (Fig. 6.3) are well established and can be found in biochemistry textbooks [3].

Upon absorption of the a first photon, the special pair D transfers an electron via the A branch to the primary quinone, Q_A. The overall timescale for this is about 200 ps. The electron is then further transferred to the secondary quinone Q_B (step A in Fig. 6.3). This reduction is closely coupled to proton transfer processes that protonate groups in the vincinity of Q_B. The timescale for the primary ET is on the order of 30-100 μs, which is significantly longer than the other electron transfer steps (see below for a discusssion of possible mechanisms).

In the next step, reduction of the special pair D is performed by cytochrome c. After absorption of a second photon and the formation of the semiquinone at the Q_A site, coupled electron-proton transfer takes place, yielding the quinolate anion Q_BH^- (step B).

A second proton transfered to Q_BH^- completes the reduction of the secondary quinone (step C). The quinole leaves the binding site in the RC into the membrane and is replenished by an oxidized quinone from the quinone pool.

The fully reduced quinole, which is soluble in the membrane due to its long lipophyllic tail (see Fig. 6.2) is oxidized by another membrane protein, cytochrome bc (Q-cycle). In the course of the oxidation the two protons bound to QH_2 are released into the periplasm (chemical proton pumping). Through the coupling of electron and proton transfer, two additional protons are pumped to the periplasmic side of the membrane (physical proton pumping). Optimally for every two photons absorbed in the photocycle, four protons are pumped through the membrane; The generated proton gradient is used to drive ATP synthesis.

While the overall reactions in the photocycle are well established, the proton transfer pathways and sequence of events in some of the steps are not clear. This work will focus on the proton transfer process coupled to the first electron transfer in step A. Here, an amino acid in close vincinity of Q_B gets protonated. This helps in stabi-

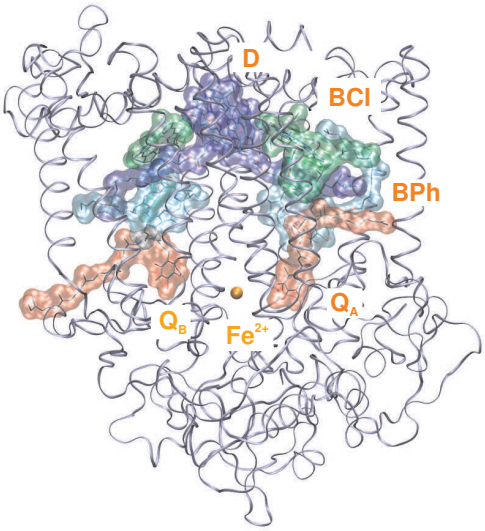


Figure 6.1: Structure of the bacterial reaction center in *Rb. sphaeroides* with cofactors: special pair of bacteriochlorophylls *D* (blue), secondary bacteriochlorophylls *BCl* (green), bacteriopheophytins *BPh* (light blue), primary quinone (red, right) and secondary quinone (red, left).

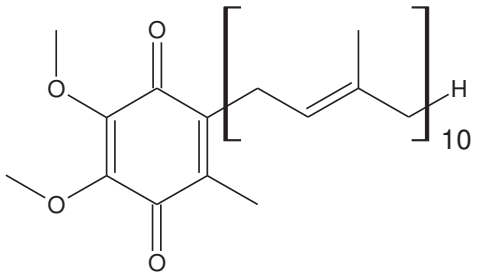


Figure 6.2: Chemical structure of ubiquinone

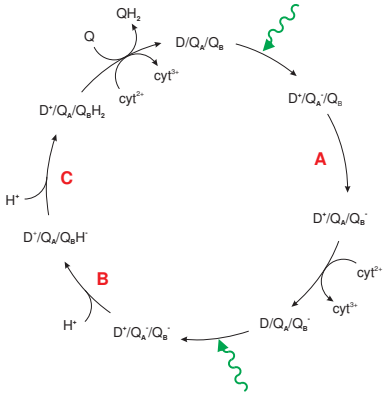


Figure 6.3: Photocycle of the bacterial reaction center. Excitation transfer or light absorption takes place in the steps marked by the green arrows. Proton transfers take place mainly in the steps denoted by A,B,C. Reduction of the special pair *D* is accomplished by cytochrome c. (cyt).

lizing the negative, radical Q_B^- after the electron transfer. The close coupling of the proton transfer and electron transfer in this step is crucial for the overall efficiency of the protein.

In Sect. 6.2 the current knowledge on the system will be reviewed, with a focus on the structural information and proton transfer reactions. Sect. 6.3 will present computations of proton transfer in step A. Sect 6.4 will conclude the chapter with a discussion of results.

6.2 Current state of knowledge: Theory and experiment

6.2.1 Structures

Among the many structures available for the BRC in *Rb. sphaeroides*, the highest resulution x-ray structure of wildtype are ones of Stowell et al. [223], with PDB entry numbers 1AIG and 1AIJ for photosystems crystallized in the dark and under ilumination respectively. The structures feature a resolution of 2.2Å and 2.6Å respectively. The light and dark adapted structures differ in the binding site of the secondary quinone, referred to as distal (dark-adapted) and proximal (light-adapted). Despite numerous experimental [224, 225] and theoretical [226–229] attempts, the role of the two binding sites has not been established. In a recent, time-resolved crystallographic study no quinone transition could be observed [225].

Further, a number of modified protein structures are available. The wildtype BRC was coordinated with Cd^{2+} for locating the proton entry (1DV3) point [230]. The light-adapted structure has a resolution of 2.5 Å. However, binding of of the heavy metal ion may perturb water structure in the PT channel. The mutant strain structure of McAuley and coworkers (1QOV) [231, 232] shows a better resolution (2.2Å) than the light adapted structure of Stowell et al. In this mutant, the exchange Ala260(M)*to*Trp displaces the primary quinone Q_A from its binding pocket. However, the perturbations in the structure are localized to the site of the mutation (Fig. 6.4). The overlay of the 1DV3 and 1QOV structures using backbone atoms shows excellent agreement with an rmsd of 0.6Å.

A number of structures exist for other site-directed mutations [233–236]. Even though some of these structures show an even higher resolution, the structural perturbations introduced by the mutations are severe.

6.2.2 Proton and electron transfer reactions

When discussing proton transfer in the BRC, two different phenomena have to be studied: proton uptake by the protein and proton redistribution within the protein. At physiological relevant pH, experimental studies observe the uptake of 0.4-0.9 protons in step A [37, 237]. A further proton is taken up either on formation of the semiquinone anion Q_BH^- [237] or shortly after full reduction of the quinone [238].

The state of knowledge is reviewed in [11, 36, 37] and more recently in [38]. A series of amino acids has been identified which participate in proton delivery to Q_B

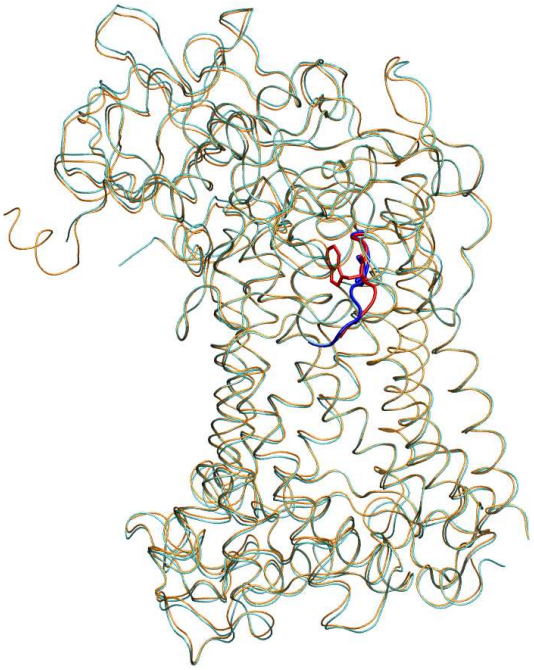


Figure 6.4: Overlay of protein structures 1AIG (light and dark blue) and 1QOV (orange and red). The area around the mutation is highlighted.

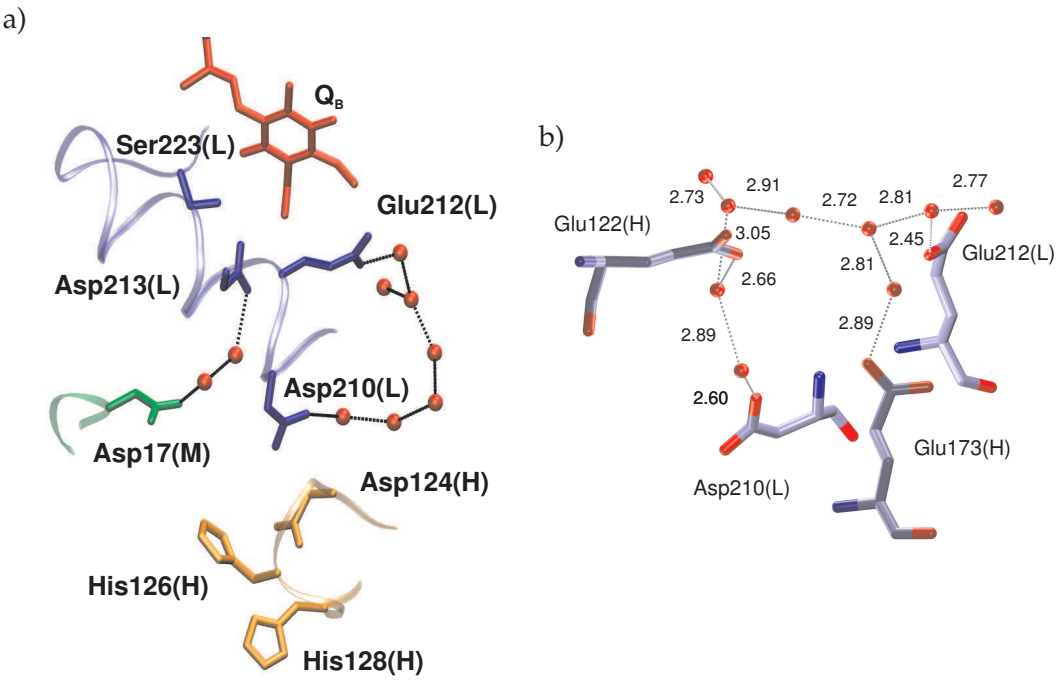


Figure 6.5: Proton transfer pathways in the BRC. a) Overview of amino acids participating in proton transfer in the bacterial reaction center. b) Water wire connecting Asp210(L) and Glu212(L). Structural information from the 1QOV structure. Only crystallographically resolved water molecules are included here.

(Fig. 6.5a).

Both protons ultimately transferred to the secondary quinone share a common entry point which is the pair of histidine residues His126(H) and His128(H) [230]. This was proven by a slowdown of the proton transfer rate when coordinating heavy metal ions such as Zn^{2+}, Cd^{2+} and Ni^{2+} to these residues. These studies were confirmed by site directed mutation and chemical rescue studies [239]. The approach to slow proton transfer using heavy metal coordination was criticized [240]. By introducing additional on the ions, the pK_a values of other nearby residues could be altered [240], hence altering the mechanism of proton transfer.

Although the main contributors to the proton transfer were identified, conclusive mechanistic predictions are involved due to the close coupling of protonable residues. Mutating Glu212(L) to Gln for instance [241] will not only eliminate this amino acid for proton transfers, but may lead Asp213(L) to take over part of the role of electrostatic modulation of proton affinities and redox potentials (vide infra).

The acidic cluster of residues responsible for proton transfer is surrounded by a large number of salt-bridges. These have been attributed to fine tuning electrostatics and possibly playing an important role for the coupling of distant residues through electrostatic domino [242]. For instance mutation of Arg233(M), despite its long distance from either quinones (ca. 15/18 Å), affects the free energy gap between $Q_A^- Q_B$ and $Q_A Q_B^-$ significantly and also alters the pH dependence of $k_{AB}^{(1)}$. The source of these changes are explained by X-Ray crystallography where large scale rearrangements of the salt bridges are observed [234].

The complexity of the problem is also reflected in a numerous theoretical studies: a number of electrostatic computations predicted the coupling of protonation and reduction states. The greatest challenges lie in the protonation states of the acidic cluster near Q_B consisting of Glu212(L), Asp213(L) and Asp210(L) and Glu173(L). For instance, the close coupling between Asp213(L) and Glu212(L) is shown by the fact that some electrostatics computations predict a constant neutral charge on Asp213(L) with proton uptake by Glu212(L) upon Q_B reduction [243, 244]. Other calculations predict switched roles [245]. In the next step, some models predict the deprotonation of Glu212(L) upon the first protonation of Q_B, whereas experimantally, the role of Asp213(L) as initial proton donor to Q_B is well established [246–248]. One common conclusion of some pK_a computations is a constant charge of -1 for the cluster Q_B, Glu212(L) and Glu213(L) in the steady state [243, 245, 249]. This underlines again the strong coupling of ET and PT.

Some discrepancies in the observation of protonation states could be due to the fact that experimental measurements are carried out for detergent solubilized RCs while calculations are performed on membrane embedded models. On the other hand, experimentally, the kinetics of the first steps of electron transfer, showed no significant difference between membrane-bound and detergent stabilized BRC [250].

These problems are also reflected on the experimental side. For instance, while time resolved experiments, such as FTIR, observe an ionized Glu212(L) [251–254] in the ground state, some older interpretations of mutation based experiments predicted it protonated [241].

Despite these inconsistencies between different measurements and theoretical pre-

dictions, a wealth of information for the processes in the photocycle is available.

Proton transfer on reduction of Q_A and proton transfer coupled to the first electron transfer to Q_B (step A)

Upon reduction of Q_A, proton uptake by the BRC takes place [237, 255]. This involves protonation of in intermediary proton donor A. Both Asp210(L) and Asp17(M) were observed to take the role of the intermediary proton donor [252–254]. The electron transfer to the secondary quinone is strongly coupled to proton transfer to Glu212(L) (step A in Fig. 6.3) which shows a high conservation in plants, algae and bacterial systems [5]. Both experimentalists [256–258] and theoreticians [243, 259] agree on the coupling of the two processes.

$$(Q_A^- Q_B) Glu^-_{212(L)} + AH \overset{k^{(1)}_{AB}}{\rightarrow} (Q_A Q_B^-) GluH_{212(L)} + A^- \tag{6.1}$$

Here, $k^{(1)}_{AB}$ is the rate constant for the overall process. Some experimental publications [241] propose that the proton transfer to Glu212(L) precedes electron transfer:

$$(Q_A^- Q_B) Glu^-_{212(L)} + AH \overset{k^{(1)}_{H^+}}{\rightarrow} (Q_A^- Q_B) GluH_{212(L)} + A^- \overset{k^{(1)}_{et}}{\rightarrow} (Q_A Q_B^-) GluH_{212(L)} + A^- \tag{6.2}$$

where $k^{(1)}_{H^+}$ and $k^{(1)}_{et}$ are the rate constants for the first PT and ET respectively.

The strong coupling of PT and ET is also shown by the fact that Glu212 (L) strongly modulates the kinetics of the first electron transfer from Q_A to Q_B depending on pH. While the wildtype reaction center shows a strong dropoff of the rate at high pH values, the Glu212(L)→Gln mutant does not show a dependence of the electron transfer rate on the pH [241]. At pH values larger than 7, $k^{(1)}_{AB}$ drops off sharply, while the rate of charge recombination of Q_A^- with the primary donor D^+ increases. This lead to the proposal, that protonation of Glu212(L) has to take place *before* the ET to Q_B. Paddock et al. estimated the pK_A of Glu212(L) in the wildtype to be around 8.5 making it an excellent proton donor to Q_B^-, $Q_B H^-$ or Q_B^{2-} [241]. Such a large pK_a shift could be explained by the hydrophobic environment of Glu212(L). The exergonicity of proton transfer from the intermittent donor to Glu212(L) is estimated to be 7.6 kcal/mol. The intermediate donor A should have a pK_a of about 4.4 [239].

Other research groups find the opposite sequence of events [252–254], i.e. proton transfer to Glu212(L) only *after* Q_B reduction:

$$(Q_A^- Q_B) Glu^-_{212(L)} + AH \overset{k^{(1)}_{et}}{\rightarrow} (Q_A Q_B^-) Glu_{212(L)} + AH \overset{k^{(1)}_{H^+}}{\rightarrow} (Q_A Q_B^-) GluH_{212(L)} + A^- \tag{6.3}$$

This was also observed for the structurally closely related reaction center of *Rhodobacter capsulatus* [257].

Remy et al. suggested a mechanism of coupled ET and PT, with an intermediate electron donor X, which was suggested to be the ferrous ion [252]:

$$\begin{aligned} Q_A^- X Q_B + Glu^-_{212(L)} + AH &\rightarrow Q_A^- X^+ Q_B^- + Glu^-_{212(L)} + AH \\ &\rightarrow Q_A X Q_B^- + GluH_{212(L)} + A^- \end{aligned} \tag{6.4}$$

Their model predicts that after reduction of Q_B by X with a time constant of 150 μs, reoxidation of X by Q_A and proton transfer to Glu212(L) occurs simultaneously$\rightarrow$, with a time constant of 1.1 ms. Further, in these experiments, deprotonation of Asp210(L) occured in the same phase as protonation of Glu212(L), identifying Asp210(L) as the proton donor. For the initial protonation of Asp210(L), a time constant of 12 μs was found. In the Asp210(L)toAsn mutant, protonation of Glu212(L) still was observed, albeit with a slower phase.

Independently, other experiments showed that in the mutant of Asp210(L)$\rightarrow$Asn the combined rate $k_{AB}^{(1)}$ of ET and PT is slowed down by a factor of 10 [253]. In mutants with both Asp210(L) and Asp17(M) replaced by Asn, a slowdown by a factor of 100 is observed. Further, sole exchange of Asp17(M) slows down $k_{AB}^{(1)}$ by a factor of 10, similar to the exchange of Asp210(L). There is evidence that also the thermodynamics of proton uptake is altered upon mutations of Asp17(M) and/or Asp210(L) [244, 256]. Zhu et al. observe from electrostatic computations that despite the close proximity of Asp210(L) and Asp17(M) both of them are ionized in the steady state [243]. Exchanging one of these residues by Asn most likely would decrease the pK_A of the other residue.

The coupled PT/ET transfer process (Eq. 6.1) is independent of the electron driving force, i.e. the difference in redox potentials of Q_A and Q_B [260]. This indicates that another process is rate limiting. The transition of the quinone between the two binding sites found by Stowell et al. (vide supra) was attributed for this gating mechanism [223]. Several experimental [223–225] and theoretical [226, 226–229] works have investigated possible gating mechanisms. involving conformational changes of the secondary quinone binding and protonation changes. Another, recent explanation for the gating mechanism is changes in the protonation state or connected to the formation of a hydrogen bond between Ser223(L) and Q_B [249].

In a recent theoretical study, Ishikata et al. discussed the influence of hydrogen bonding from Ser223(L) to the secondary quinone on redox potential of the latter [249]. In their study they observed a hydrogen bond from Ser223(L) to the negatively charged semiquinone Q_B^- only if Asp213(L) had a charge corresponding to a 75% protonation. This hydrogen bond is of essential importance for transferring the first proton to Q_B (see below). Its presence was recently observed using ENDOR spectroscopy [261] for the semiquinone radical.

Proton transfer coupled to the second electron transfer to Q_B (step B)

The second electron transfer is highly coupled to proton transfer to Q_B:

$$(Q_A^- Q_B^-) + H^+ \overset{k_{AB}^{(2)}}{\rightarrow} (Q_A Q_B H^-) \tag{6.5}$$

Here, $k_{AB}^{(2)}$ is the rate constant for the coupled PT/ET. There is experimantal evidence, that here, proton transfer precedes proton transfer:

$$(Q_A^- Q_B^-) + H^+ \overset{k_{H^+}^{(2)}}{\rightarrow} (Q_A^- Q_B H) \overset{k_{et}^{(2)}}{\rightarrow} (Q_A Q_B H^-) \tag{6.6}$$

where $k_{H^+}^{(2)}$ and $k_{et}^{(2)}$ are the respective rate constants for PT and ET in this step. The protonated semiquinone was observed for BRCs in which the ubiquinone was exchanged for a more basic rhodoquinone [262]. Essential residues for the proton coupled electron transfer in the species *Rb. sphaeroides* are Ser223(L) and Asp213(L), with the latter being the proton donor. Ser223(L) serves as the bridge between Asp213(L) and Q_B.

Replacement of Asp213(L) with Asn decreases the rate of proton transfer by a factor of about 10^6 leading to a standstill in the photosynthetic cycle [246, 247]. Through second site mutation studies the proton transfer was restored; the mutation Asn44(M)*to*Asp introduces an alternate proton donor [248]. This double mutant replicates the structure of the reaction center in *Rps. viridis* in which has both Asp213(L) substituted by an Asn and Asn44(M) by Asp.

Paddock et al. proposed that the role of Asp213(L) is twofold by both transferring the first proton to Q_B^- and by stabilizing the proton on the resulting intermediary radical Q_BH [247]. The Ser223(L)→Ala mutant shows a slowed electron transfer rate $k_{AB}^{(2)}$ by a factor of 350, measured by UV-Vis spectroscopy [263].

Second protonation of Q_B (step C)

Beyond modulating the primary electron transfer to Q_B (vide supra), Glu212(L) also donates the second proton to Q_B. This essential role was shown in a drastically reduced cytochrome turnover after the first reoxidation for the mutated RC [241] which was shown to have a well preserved structure [233]. This indicates the importance of Glu212(L) as a proton donor:

$$(Q_A Q_B H^-) GluH_{212(L)} \rightarrow (Q_A Q_B H_2) Glu^-_{212(L)} \tag{6.7}$$

The exact pathway of the proton transfer is yet unknown.

6.3 Proton transfer from Asp210(L) to Glu212(L)

In the following, results for computations of the proton transfer from Asp210(L) to Glu212(L) (step A) through the channel depicted in Fig. 6.5 are presented. The goal is to investigate whether from the point of proton transfer kinetics, proton transfer occurs before or after electron transfer.

6.3.1 Computational details

The coordinates for the simulations were taken from the structure of McAuley et al. (1QOV) [231, 232]. The positions of the primary quinone Q_A and the residues 257-261 were reconstituted from 1DV3.

Protonation states for the ionizable residues were adapted from electrostatic computations of Ullmann et al. (G. M. Ullmann, private communication). All amino acids were kept in their default protonation state, except for Glu104(L), Asp210(L), Asp213(L), His68(H), His126(H), His141(H), which were protonated.

For the simulations, the origin was centered on the carbonyl carbon atom in Pro209(L) and oriented such that the membrane normal points in z-direction. A cavity of 23 Å centered on the origin was defined for the GSBP/stochastic boundary conditions protocol. Heavy atoms outside a radius of 18 Å were harmonically constrained and treated using Langevin dynamics. To include bulk solvent, the sturcture was overlayed with a 21 Å sphere of water molecules, which were restrained using an MMFP potential. All molecular dynamics simulations were carried out coupled to a heat reservoir at 300 K using Langevin dynamics in a stochastic boundary condition setup as described in the literature [127]. For the MM simulations, the X-H distances were kept constant using SHAKE [140] and a timestep of 2 fs was chosen.

For the GSBP computations, a dielectric constant of 80 was chosen for the solvent. The membrane dielectric constant was 2, with a 40 Å membrane thickness. Spherical harmonics up through the 20th order were used as the basis functions. For comparison, in addition to the membrane embedded protein (membrane model), simulations of the proton transfer process where also carried out for an entirely solvated system lacking the membrane (solvent model).

For the QM/MM simulations, the QM zone comprised 48 atoms, and includes the sidechains of Asp210(L) and Glu212(L) and the 9 water molecules highlighted in Fig. 6.7.

For the umbrella sampling simulations the modified center of excess charge coordinate was used in conjunction with ζ_R (Eq. 5.11) as the reaction coordinate. Atom labels and weights are given in Fig. 6.6. As in Glu212(L), either of the carboxyl oxygen atoms (Glu212(L):$O^{\epsilon 1}$,Glu212(L):$O^{\epsilon 2}$) can accept the proton ζ' was included for symmetrization. Although not needed in the case of Asp210(L) (Asp210(L):$O^{\delta 1}$,Asp210(L):$O^{\delta 2}$), the corresponding term ζ' was added for balance. As coordinate for donor and acceptor, the center between the respective carboxyl carbon atom and oxygen atoms was taken. Umbrella sampling involved the simulation of 40 independent windows with force constants ranging from 0-500 kcal/mol. The simulations each consisted of 20 ps of heating and equilibrattion, and a further 50 ps for production and data accumulation.

For the excess coordination analysis, the expression introduced by Chakrabarti et al. [17] was used to determine the number of excess protons coordinated to each oxygen atom (Eq. 5.3). As the ζ_R coordinate is non-linear with respect to distance from either the initial donor or the final acceptor, a similar scalar expression for determining the oxygen positions $\zeta_O(O_i)$ is used, using the distances r_{DO_i} and r_{AO_i} from the donor and acceptor to the respective oxygen atom:

$$\zeta_O(O_i) = \frac{r_{DO_i} - r_{AO_i}}{r_{DO_i} + r_{AO_i}} \tag{6.8}$$

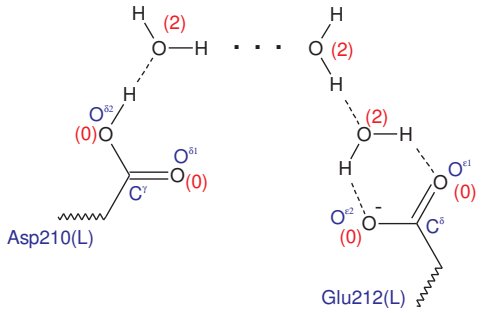


Figure 6.6: Atom labels and weights (in parentheses) associated with the definition of the modified center of excess charge coordinate from Asp210(L) to Glu 212(L) in the bacterial reaction center

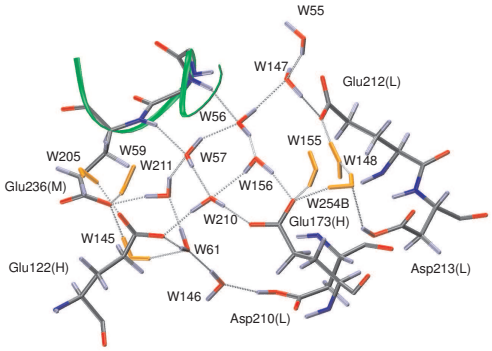


Figure 6.7: Structure of the channel obtained from MD simulation. Water molecules described as MM in the QM/MM simulations are rendered in orange.

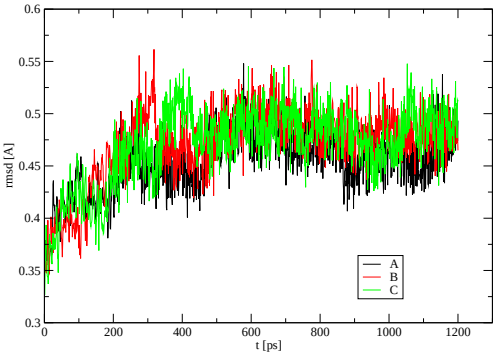


Figure 6.8: Rmsds for backbone atoms in the inner cavity for three molecular dynamics simulations in the $Q_A^- Fe^{2+} Q_B^0$ charge state of the BRC.

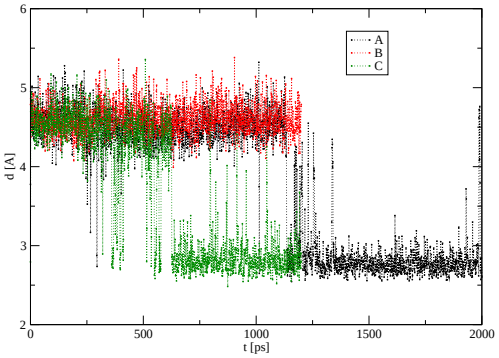
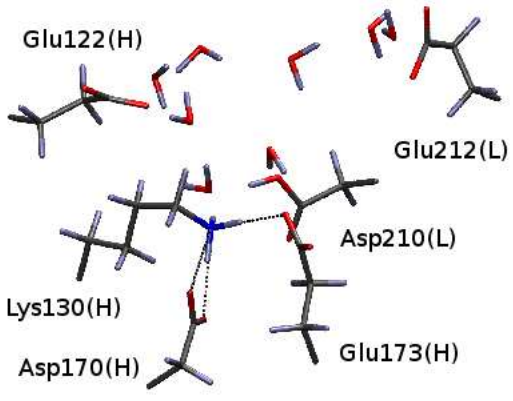


Figure 6.9: Evolution of the distance between Lys130(H):N^Z and Glu122(H):$O^{\epsilon 1}$

a) b)



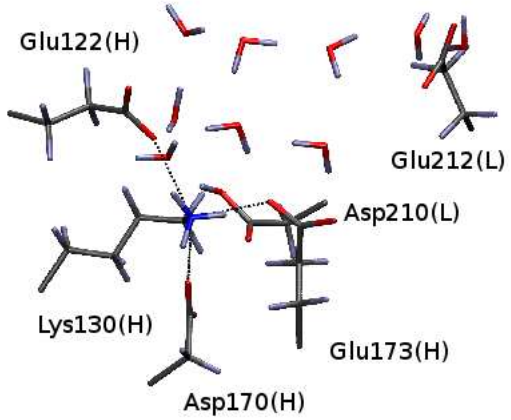


Figure 6.10: Snapshots displaying the structure around Lys130(H) a) before and b) after formation of the salt bridge to Glu122(H).

6.3.2 Molecular Dynamics Simulations for the reactant

Molecular mechanics based MD simulations were run for the $Q_A^- Fe^{2+} Q_B^0$ charge state to obtain a reasonably representative starting point for the PMF simulations. The convergence of the RMSDs for the backbone atoms in three MD simulations are shown in Fig. 6.8. The convergence of the rmsds show that on the timescale considered here, the simulations are equilibrated.

In two of the three simulations examined here, an additional salt bridge is formed close to the channel. In the initial structures, Lys130(H) forms a hydrogen bond with Asp170(H) and Glu173(H) (Fig. 6.10a). Later, an additional hydrogen bond to Glu122(H) can be found (Fig. 6.10a). The time series of distances between the nitrogen atom of the lysine and the carboxyl oxygen atom of Glu122(H) show that this salt bridge does not form immediately (Fig. 6.9). However, the longer simulation (A) shows that the salt-bridge is stable over a longer period of time.

This salt-bridge is not modelled current X-ray structures (e.g., 1PCR, 1AIJ, 1AIG, 1DV3, 1QOV). The structural changes involved with the formation of this salt-bridge are minor (Fig. 6.10), however the salt-bridge also plays an important role for the water structure between Asp210(L) and Glu212(L): an additional water molecule has diffused into the space between Glu122(H) and Glu173(H). As we will see later in the analysis of the PMF results, this water molecule plays an important role in proton transfer.

As an interesting detail, note the water molecules W148 and W254B in Fig. 6.7, which have diffused into the space between Asp213(L) and Glu212(L) during the molecular dynamics simulation. These water molecules could be part of the pathway for Glu212(L) protonation in the Asp210(L)→Asn mutant. However, more rigorous treatment is needed for determining, whether these positions are thermodynamically stable, e.g. grand-canonical Monte-Carlo [264].

6.3.3 Potential of mean force computations

Here, we investigate the two charge states $Q_A^- Fe^{2+} Q_B^0$ and $Q_A^0 Fe^{2+} Q_B^-$. We further include $Q_A^- Fe^{3+} Q_B^-$, a structure suggested for an intermediate in the electron transfer, which was recently suggested by Remy et al. [252]. The computations were carried out both for a model of the membrane embedded protein, as well as a solvent embedded protein.

The potential of mean force computations were started out from the structure depicted in Fig. 6.7, which was discussed above. The PMFs are depicted in Fig. 6.11 and energies of critical states relative to the reactant are listed in Tab. 6.1.

The PMFs for the different charge states and models all feature a transition state at $\zeta_R \approx$ -0.25. The energies of these states relative to the respective reactants vary between 14.7 and 16.3 kcal/mol. The PMFs for the $Q_A^- Fe^{2+} Q_B^0$ charge state feature a second maximum at $\zeta_R \approx 0.20$ and 0.25 for the membrane and water model respectively. The energies of these second maxima are comparable to the first maximum. For the charge states where Q_B is negatively charged, the barriers is located at $\zeta_R \approx$ -0.25 and is slightly lower. The occurrence of the second maximum in the $Q_A^- Fe^{2+} Q_B^0$

Table 6.1: Exergonicities and barrier heights (in kcal/mol) obtained from the PMFs.

	$W^{\ddagger}$	ΔW
membrane model		
$Q_A^0 Fe^{2+} Q_B^0$	16.5/18.6	-2.7
$Q_A^- Fe^{2+} Q_B^0$	16.3/17.5	-3.1
$Q_A^- Fe^{3+} Q_B^-$	15.2	-11.1
$Q_A^0 Fe^{2+} Q_B^-$	16.1	-15.6
solvent model		
$Q_A^- Fe^{2+} Q_B^0$	15.7/15.5	-5.4
$Q_A^0 Fe^{2+} Q_B^-$	14.7	-17.3

charge state can be understood in terms of the Hammond postulate, taking into account that the exergonicity is smaller than for the other charge states. No substantial difference in the barrier heights can be found from this model which allow a conclusion of whether proton transfer is more favorable in one charge state.

Both of the quinones are close to the surface of the protein. Due to the long-range interactions involved, e.g. the distance between Glu212(L) and Q_A is about 15 Å, one would expect, that solvent shielding plays an essential role for the energetics of this PT. It is surprising to see only modest changes of both barriers and exergonicities. However, these results are in agreement with experiments of Beekman et al. [250].

The exergonicities vary greatly. The large exergonicity where Q_B is negatively charged (-11.1, -15.6 and 17.3 kcal/mol), can be understood from the close proximity of Glu212(L) to Q_B (see Fig. 6.5a). For this case, the proximity of the negative charge on Q_B and Glu212(L) leads to a destabilization of the reactant state.

The further analysis will focus on the computations performed for the $Q_A^- Fe^{2+} Q_B^0$ charge state in the membrane model. The analysis uses snapshots taken every 100 fs of the MD trajectories.

The trace of the location of the excess proton computed using the mCEC (Fig. 6.12a) allows us to gain some insight into the reaction. Here, for each of the snapshots from the trajectories, the mCEC was computed and entered as a datapoint. One clear testimony from this trace is that the water wire here is highly curved. Further, ζ_R permits the system enough flexibility to explore the two branches of the proton wire. For instance in the region $-0.2 < \zeta_R < 0.2$ (Fig. 6.12c), the width of the distribution orthogonal to the path clearly indicates participation of both branches. Examining the mCEC vectors for different values of ζ_R, we can locate the transition states at $\zeta_R = -0.2$ and $\zeta_R = 0.25$ near W61 and W57. Due to the incomplete data set used here and the fact that the individual trajectories are biased by the umbrella potential, the *densities* of data points in three dimensional space do not allow a conclusion about the relative energetics.

For a more quantitative analysis of the mechanism of proton transfer, we can use excess coordination analysis. It allows the identification of the protonated water species participating in proton transfer and hence allows the distinction between synchronous and asynchroneous PT. A synchronous mechanism involves simulta-

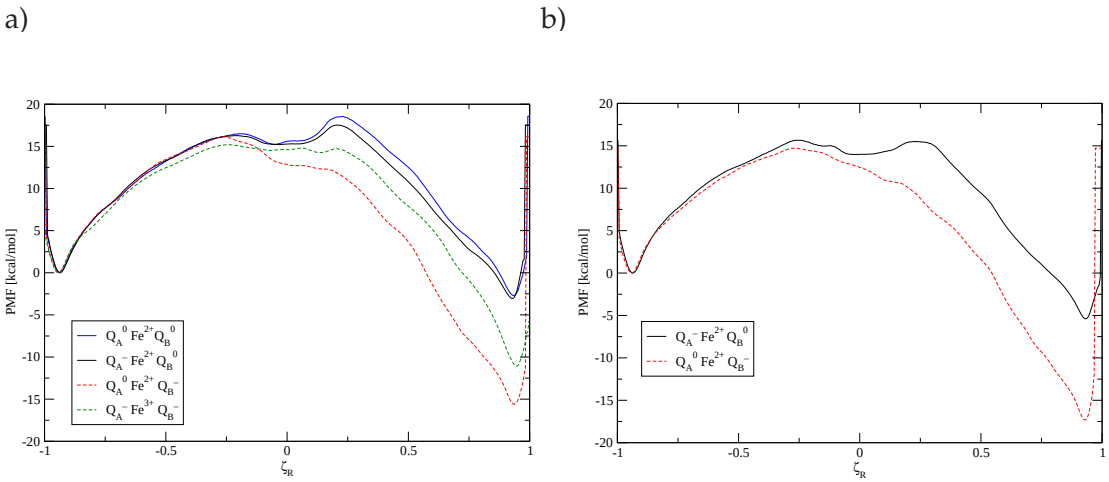


Figure 6.11: PMF for the proton transfer from Asp210(L) to Glu212(L) in the a) membrane model and the b) solvent model.

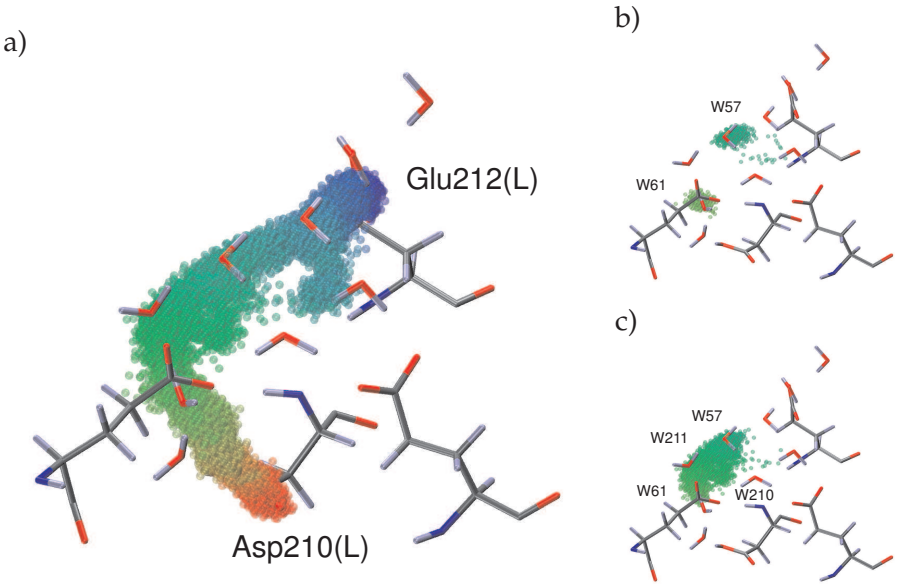


Figure 6.12: Trace of the location of the excess proton computed by mCEC. The color encodes the value of the ζ_R coordinate from -1 to 1 (red-green-blue). a) all snapshots over the entire range of the reaction coordinate b) snapshots for the transition states with $-0.27 < \zeta_R < -0.23$ or $0.17 < \zeta_R < 0.23$ c) intermittent snapshots with $-0.2 < \zeta_R < 0.2$.

neous proton transfer over a longer chain of water molecules and hence change of hydrogen coordination for all oxygen atoms involved. This includes extended protonated clusters, i.e. Zundel ions oriented in direction of the pathway. In contrast, an asynchroneous mechanism implies defined smaller protonated clusters, i.e. hydronium ions possibly in the form of Eigen-ions.

In Fig. 6.13 the excess coordination is shown in two different representations. Each of the peaks in the plot represents a protonated water molecule or a protonated amino acid. In the reactant state, only Asp210(L) is protonated, while in the product state Glu212(L) is protonated. For the proton transfer studied here, we observe different phases of synchroneous and asynchroneous mechanism (Fig. 6.13). In regions A and C, we see a series of three peaks oriented parallel to the axis of oxygen positions. For these areas, we observe multiple water molecules with more than two protons coordinated to them, with a simultaneous residual coordination to the neighboring carboxylic acids. To illustrate the structures involed, snapshots from the windows $\zeta_R = -0.4, -0.20$ are shown in Fig. 6.14b) and e). Thus, two water molecules are directly involved in protonation and deprotonation of the carboxylic acids at both ends.

In contrast, in region B from about ζ_R=-0.2 to 0.2, an asynchroneous mechanism takes place. Here, the peaks indicating protonated water molecules are oriented along diagonal axis, i.e. for one value of the reaction coordinate, mostly one oxygen atom shows excess protonation. The mechanism here is hence sequential protonation and deprotonation. The corresponding snapshots are shown in Fig. 6.14c)d).

The change of character in the mechanism observed here can be understood in terms of the strong curvature of the wire in this area (Fig. 6.12a) and c). The role of the two glutamates Glu122(H) and Glu236(M) limiting the channel in this area is to stabilize the more isolated ions observed here. Further, the MM water molecules around Glu122(H) – W145, W205, W211 shown in Fig. 6.7 can participate in the formation of Eigen-ions. In contrast, the ion structures found in region A and C do not show this possibility of solvation.

The transition states of the reaction at $\zeta_R = -0.25$ to $\zeta_R = 0.25$ coincide with the change of mechanism between regions A/B and B/C.

To clarify some of the aspects of the mCEC trace and the excess coordination analysis, representative snapshots from different windows are shown in Fig. 6.14. The reactant and product are shown in Fig. 6.14a) and f). A snapshot close to the transition state from the $\zeta_{R0} = -0.2$ window depicts a hydronium ion at W61. This is in agreement with the analysis of the mCEC trace in Fig. 6.12b). This ion is hydrogen bonded to both W146 and W211, indicating a continuation of proton transfer across W211. The trace of the mCEC confirms that this is one possible pathway (Fig. 6.12c). However, the broadness of the distribution of states orthogonal to the path suggests that another parallel pathway across W210 possible. A realization of this pathway is depicted in snapshot Fig 6.14d). The snapshot of a transition state related structure in Fig. 6.14e) confirms the existence of extended protonated water clusters participating in synchronous water transfer as the analysis of the bond orders showed.

During the proton transfer, the solvation of Asp210(L) changes. While in the reactant state only one water molecule is coordinated to side facing the cytosol (Fig. 6.14a)-c), two water molecules are coordinated after the first transition state (Fig. 6.14d)-

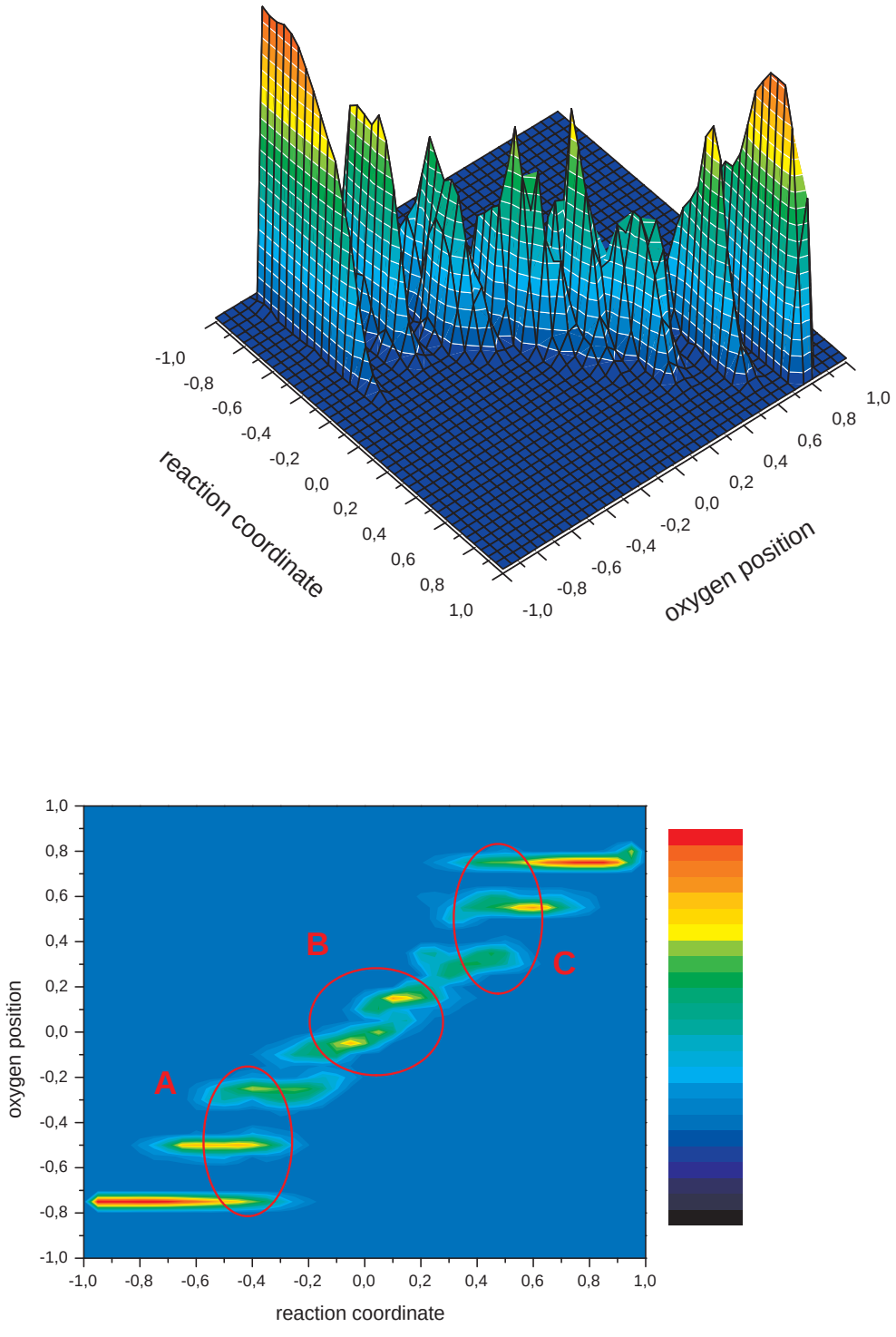


Figure 6.13: Bond order analysis for proton transfer for snapshots taken from the umbrella sampling simulations: The number of excess protons coordinated to each oxygen atom (Eq. 5.3) is shown. The reaction coordinate is ζ_R, the oxygen position is determined using $\zeta_O(O_i)$ Eq. 6.8. The coordinating oxygen atoms of the amino acids Asp210(L) and Glu212(L) are located at $\zeta_O \approx$-0.8 and 0.8. The lower panel shows the projection into the X-Y plane.

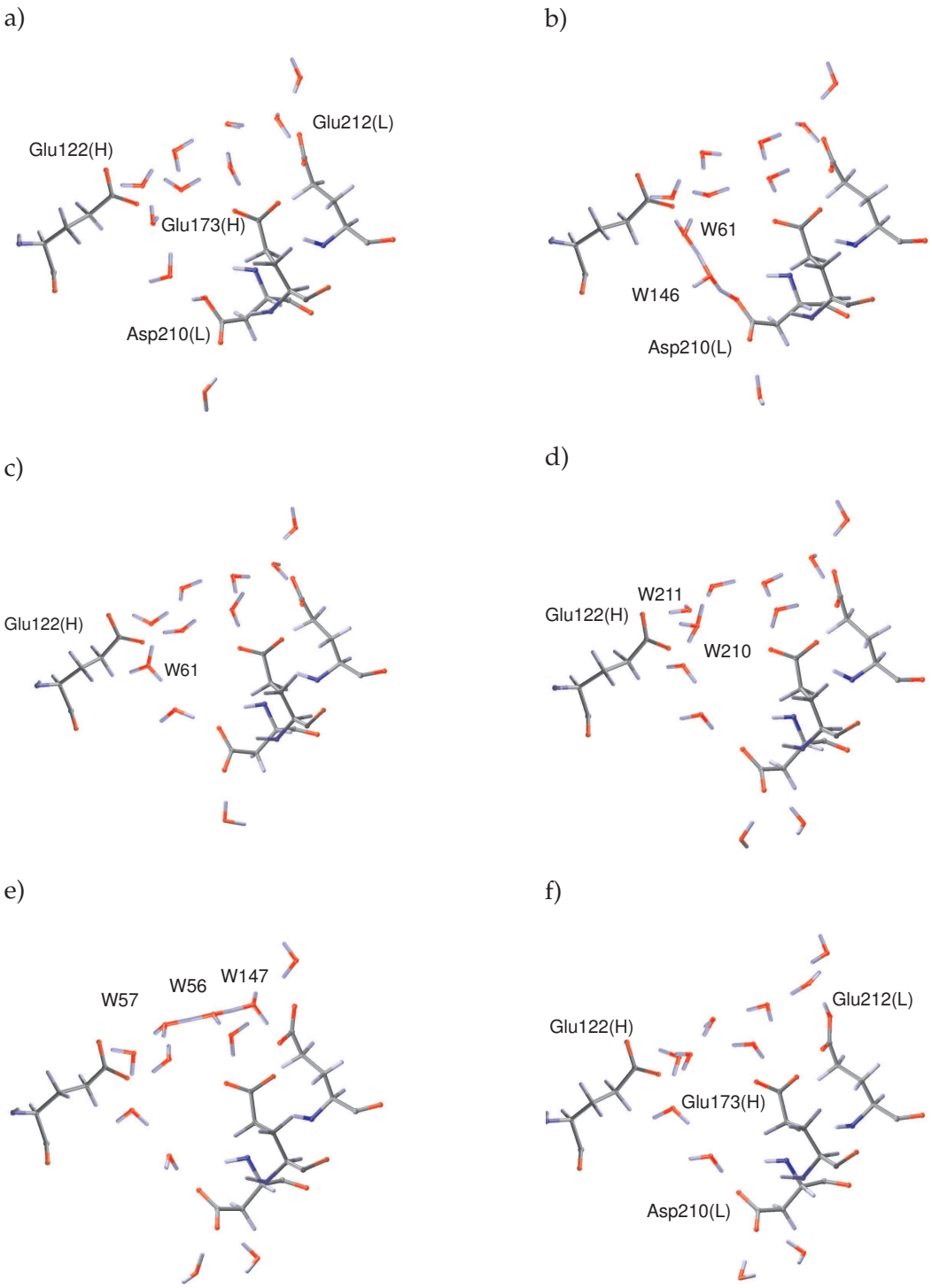


Figure 6.14: Selected snapshots of umbrella sampling simulations of the $Q_A^- Fe^{2+} Q_B$ charge state. Windows with ζ_r= a) -0.95 b) -0.40 c) -0.20 d) 0.00 e) 0.20 f) 0.95. The distance cutoff between the oxygen atoms and hydrogen atoms was chosen as 1.3 Å.

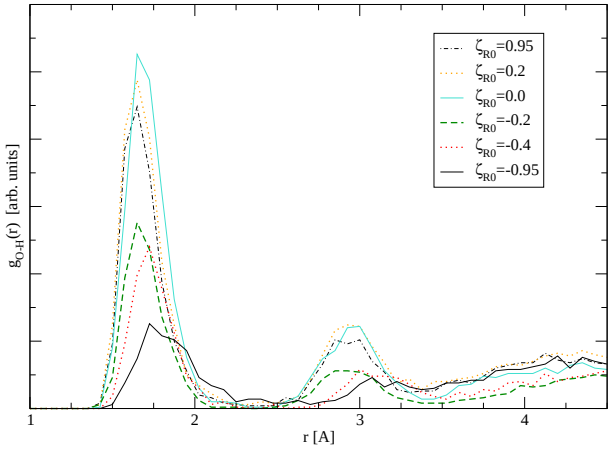


Figure 6.15: Radial distribution function g_{HO} for non-QM water hydrogens around the carboxylic oxygens of Asp210(L):O^{δ}, plotted for different windows of umbrella sampling simulations with the harmonic potential centered at ζ_{R0}.

f). This change in solvation can also be seen in the radial distribution functions (Fig. 6.15). In the reactant around the neutral aspartic acid, the first solvation shell is very broad. With increasing distance of the reaction coordinate, transitioning towards aspartic acid, the first solvation shell gets more populated. Further a second solvation shell develops around 2.9 Å.

6.4 Discussion and Conclusion

The barriers found here are slightly higher than predicted by experimental measurements [252]. The reason could be found in the classical treatment of hydrogen nuclei, neglecting zero-point energy and tunneling effects. Further, benchmark calculations show that SCC-DFTB slightly overestimates the barriers for proton transfer in similar situations (Appendix A)

From the PMFs computed here, we cannot deduce that proton transfer in one of the charge states is more favorable. A number of reasons can be considered as explanation:

1. The approach here was to investigate the proton transfer starting out the molecular dynamics simulations from identical snapshots of a long molecular dynamics run of the $Q_A^- Fe^{2+} Q_B^0$ charge state.

   However, this does not mean, that in the protein shows no preference for proton transfer in either charge state. We could only show that the mechanism for controlling the proton transfer is not merely the electrostatic interaction with Q_A or Q_B.

   Rather, the mechanism controlling proton transfer might be grounded in the electrostatic domino of the numerous salt bridges around the channel [234, 242].

Electrostatic domino might also be one way of explaining other long-range relations. For instance, proton uptake by Asp210(L) in step A is triggered by electron transfer to Q_A [252, 254], and in fact also takes place in Q_B depleted BRC [254].

2. Polarization of the water wire might play an important role. The QM/MM approach used here, allows for polarization of the QM zone by the MM atoms, but not vice versa. The size of the QM zone can hence influence the energetics. Polarization and cooperative effects were found to be important issues for the computation of proton affinities in water clusters [265]. Inclusion of further water molecules into the QM zone, notably those around W61 and W211, might lower the barrier at $\zeta_R = -0.2$, by stabilizing protonated clusters in this area.

   However, the respective water molecules W205, W145 and W59 show high mobility even on the timescale of the umbrella sampling simulations (data not shown), being replaced by other MM water molecules in the course of the simulation. This makes the inclusion of these water molecules into the QM zone problematic, and the desired effect, stabilization by cooperative effects, cannot be obtained. Here, strategies to dynamically exchange the QM and MM selection could provide a solution [266].

   Further, electronic polarization of the two glutamate residues in this area, Glu211(H) and Glu236(M) might play an important role.

   While increasing the size of the QM zone in this area could lower the barrier for proton transfer, also a shift in the importance of the two possible barriers seen in different protonation states could occur.

3. One, trivial explanation could be that there is no clear preference for the proton transfer in either charge state from a kinetic point of view. Hence, the proton transfer can take place in either charge states. However, electron transfer preceding proton transfer might be unfavorable enough to not occur beforehand. The change in exergonicities for different Q_B charge states confirms the strong coupling between PT and ET and that indeed from the point of electron transfer, electron transfer should occur *after* proton transfer.

The bond order analysis shows that during the proton transfer a change in the mechanism occurs. This change in mechanism can be attributed to the kink in the proton wire. The transition states coincide with the change in mechanism.

Further the results show that umbrella sampling simulations with ζ_R implicitly describe solvent reorganization. While description of solvation is not a part of the reaction coordinate, we see a continuous adaption of the solvation in the umbrella sampling simulations, stabilizing the respective charge state. This shows the importance of a dynamic approach to long-range proton transfer, as in adiabatic mapping simulations, the solvent molecules cannot overcome the (small) barriers for reorganization. The importance of dynamic solvent orientation for each step in proton transfer was also stressed by other authors [22, 40]. The observed solvent reorientation stresses the importance of dynamical approaches to long-range proton transfer.

Chapter 7

Summary and conclusion

The main contributions of this work are of methodological developments for the description of long-range proton transfer and their application to the bacterial reaction center.

Assessment of QM/MM frontier treatments

In Chap. 4 the importance of the QM/MM frontier treatment was discussed. Different schemes for the QM/MM electrostatic treatment in the CHARMM/SCC-DFTB framework were implemented and their performance was evaluated. In addition to the analysis of gas-phase molecules as in previous studies, an important element was the examination of chemical reactions in realistic enzyme systems.

Different schemes for SCC-DFTB/CHARMM simulations were tested, which treat electrostatic interactions due to the MM atoms close to the QM/MM boundary in different ways. Further, a new approach, the *divided frontier charge (DIV)*, was implemented in which the partial charge associated with the frontier MM atom (link host) is evenly distributed to the other MM atoms in the same group. The performance of these schemes was evaluated based on properties including proton affinities, deprotonation energies, dipole moments, and energetics of proton transfer reactions. Similar to previous work, it was found that calculated proton affinities and deprotonation energies of alcohols, carbonic acids, amino acids, and model DNA bases are very sensitive to the link atom scheme. The commonly used single link atom approach often gives error on the order of 15 to 20 kcal/mol. Other schemes give better and, on average, mutually comparable results. For proton transfer reactions, encouragingly, both activation barriers and reaction energies are fairly insensitive (within a typical range of 2-4 kcal/mol) to the link atom scheme due to error cancellation, and this was observed for both gas-phase and enzyme systems.

divided frontier charge (DIV)

Development of new reaction coordinates for proton transfer

The development of a new set of reaction coordinates was described in Chap. 5. In contrast to earlier suggestions, the *modified center of excess charge (mCEC)* allows to precisely locate the excess proton in the water chain and eliminate some problems

modified center of excess charge (mCEC)

with respect to water orientation and water collisions. This allows the computation of the potential of mean force (PMF) in linear water chains as demonstrated in a joint publication [124].

Using the mCEC, a set of collective reaction coordinates was introduced, which describes the translocation of the mCEC between the initial proton donor and final acceptor. This allows the treatment of proton transfer along nonlinear and branched pathways. The resulting reaction coordinates suggested here, ζ *and* ζ_R were shown to be applicable for adiabatic mapping and for the computations of the PMF.

ζ and ζ_R

The new coordinates ζ and ζ_R do not *a priori* assume a sequence of steps for proton transfer as is the case with previous alternatives, e.g., the asymmetric stretch coordinate. Furthermore, as the precise pathway is not predetermined, ζ and ζ_R are optimum candidates for systems with fluctuating water wires.

The new techniques introduced by the mCEC, ζ and ζ_R are already in use for other projects by a number collaborators: For the description of proton diffusion and autoionization in aqueous solutions (Y. Yang, X. Prat Resina, H. Yu, Q. Cui et al.) and for the description of long-range proton-transfer in other proteins such as aquaporin (M. Hoffmann, Th. Frauenheim et. al.), cytochrome C oxidase (N. Ghosh, X. Prat Resina and Q. Cui) and carboxylic anhydrase II (D. Riccardi and Q. Cui).

Proton transfer in the bacterial reaction center

As an application of the techniques developed here, proton transfer along a highly curved, ∼20 Å long pathway in the bacterial reaction center (BRC) was computed. The residues Asp210(L) and Glu212(L) are connected by a bifilar chain of eight water molecules and are 10 Å apart.

One clear proof from these calculations is that ζ_R can describe proton transfer in a highly curved, *S-shaped* water wire. Further, ζ_R permits the system enough flexibility to explore either pathways in the branched wire. Depending on the charge state of the quinones, the barrier for PT is 15-17 kcal/mol. This is only slightly higher than what is expected from experiment. For the Q_B^- charge states, the transition state is earlier than the highest transition state found for the Q_B^0 charge states. Despite these phenomenological differences, the results for this model suggest that the sequence of events for proton transfer cannot solely be understood from direct electrostatic interaction with the quinones. Instead, the foundation may lay in an electrostatic domino effect, i.e. the rearrangement of salt-bridges between Q_A and the proton wire. In addition, the results show that the solvent structure changes significantly during the reaction, stressing again the importance of dynamical approaches to long-range proton transfer.

Outlook

In this work one part of the proton transfer processes in the bacterial reaction center was described. Even though this work addressed some important questions, many steps and details of the proton transfer remain to be clarified, e.g.:

- For the proton transfer to Glu212(L), the precise mechanism of protonation of the intermediate donor Asp210(L) is yet unknown.
- In bacterial reaction centers in which Asp210(L) was mutated to Asn an alternate pathway using Asp17(M) as the intermediate donor is possible.

These reactions share the property that they convey protons through a channel with highly mobile water molecules. The reaction coordinate introduced in this work was developed with situations like this in mind for describing proton transfer through alternating channels. However, the simulations are currently limited by the description of the water molecules diffusing away. A suitable protocol for flexibly determining the QM/MM selection is in development by other researchers and will help overcome this limitation [266].

Further, the second proton transfer to Q_B before the fully reduced quinone leaves the binding pocket, is not yet understood. While the sequence of events in the cycle are clear and the proton donor was identified as Glu212(L), the pathway is not clear from the protein structure. Despite a number of propositions made for the mechanism, no studies have able to shed light on this process.

Some of these riddles of proton transfer in the bacterial reaction center will be attacked in the future using the methods and protocols developed here.

Conclusion

In this thesis further steps towards the understanding of long-range proton transfer were made. Beyond the initial goal of describing proton transfer in the bacterial reaction center, the developments lay the groundwork for exploring proton transfer in other biologically and non-biologically relevant systems.

Appendix A

Benchmark computations

Proton affinities

Proton affinities for relevant molecules for the proton transfer process in the bacterial reaction center studied in this work are shown in Tab. A.1. The geometry of all molecules was optimized at the respective level of theory. Ab-initio and DFT computations were carried out using Gaussian98 [267].

Table A.1: Proton affinities for relevant molecules in kcal/mol.[a]

	H_2O	$(H_2O)_2$	Ac^-
SCC-DFTB	-180.4	-207.0	-366.4
B3LYP/6–311++G(2d2p)	-170.6	-201.0	-353.8
BP86/6–311++G2d2p	-170.7	-202.5	-352.0
CCSD/cc–pVTZ	-172.3	-200.7	-357.3
CCSD(T)/cc–pVTZ	-171.5	-200.4	-355.3

[a] Ac^- = acetate,AcH= acetic acid

The results show, that SCC-DFTB consistently overestimates proton affinities for the molecules presented here. The relative proton affinities however, are better reproduced.

Proton transfer

For evaluating the reliability of barriers for proton transfer, the system AcH-$(H_2O)_n$-Ac^- (see footnote in Tab. A.1) was chosen, were n is the number of bridging water molecules. The geometry for this testcase was chosen such that all of the carboxylic oxygen and carbon atoms are constrained the same plane, with the connecting O-O vectors in both molecules being parallel. The distance d between the respective carboxylic oygen atoms was varied and proton transfer studied using the ζ_R coordinate (Chap. 5). A force constant of 2000 kcal/mol was applied to ζ_R during the mapping in each step. The fully self consistent calculations were carried out at geometries optimized at the SCC-DFTB level.

The results are shown in Fig. A.1-A.4. The contour plots show the energies along the proton transfer, relative to the energy of the reactand at ζ_R=-1.0 for the same donor

acceptor separation d.

While the *absolute* proton affinities are consistently overestimated, the barriers for proton transfer are well reproduced, in comparison to the B3LYP computations show that the barriers for proton transfer are well reproduced. For average values of d, SCC-DFTB overestimates the barrier by about 2-3 kcal/mol. For large donor-acceptor distances, the deviations are lower than 5 kcal/mol.

However, the energetics for larger donor-acceptor distances d are not representative of what can be found in realistic proton wires. Evidence for this can be found in assymetries in the energetics, e.g. at d=13.2-13.4 in the case of a four water-molecule wire (Fig. A.4). Here, slight discontinuities in the energy can be found. This means that the coordinate rearrangements can no longer be described using ζ_R alone. In fact, here, for proton transfer to occur, the water oxygen atoms have to significantly change positions during proton transfer.

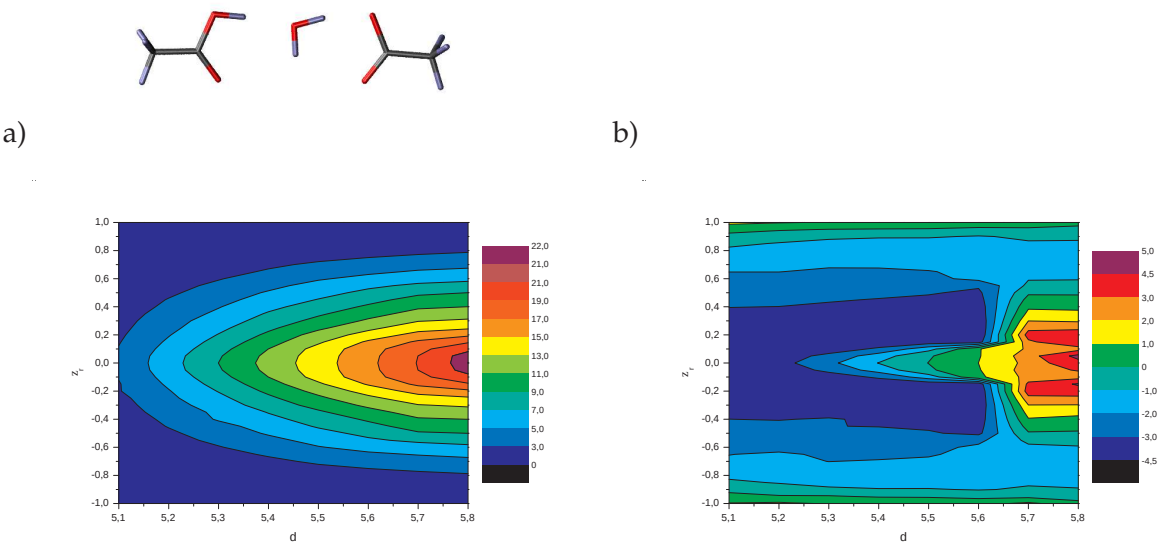


Figure A.1: Benchmark results for a one-water bridge between acetic acid and acetate. Energies (in kcal/mol) are relative to the energy of the reactand at ζ_R=-1.0 for the same value of d. a) SCC-DFTB energies along the reaction path for different donor-acceptor distances d, c) B3LYP/6-311G** - SCC-DFTB results.

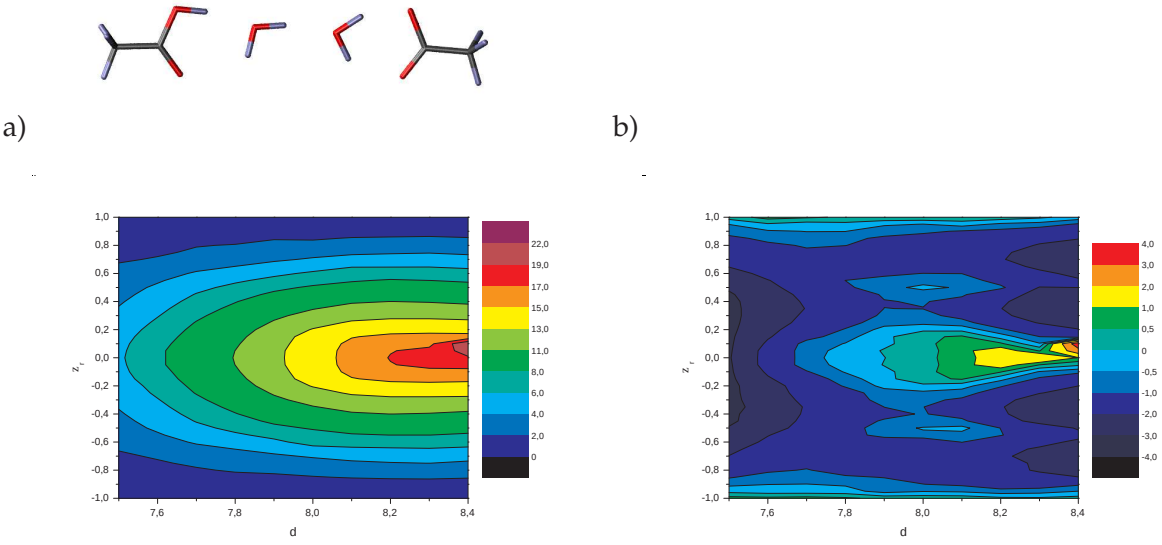


Figure A.2: Benchmark results for a two-water bridge between acetic acid and acetate. Energies (in kcal/mol) are relative to the energy of the reactand at ζ_R=-1.0 for the same value of d. a) SCC-DFTB energies along the reaction path for different donor-acceptor distances d, c) B3LYP/6-311G** - SCC-DFTB results.

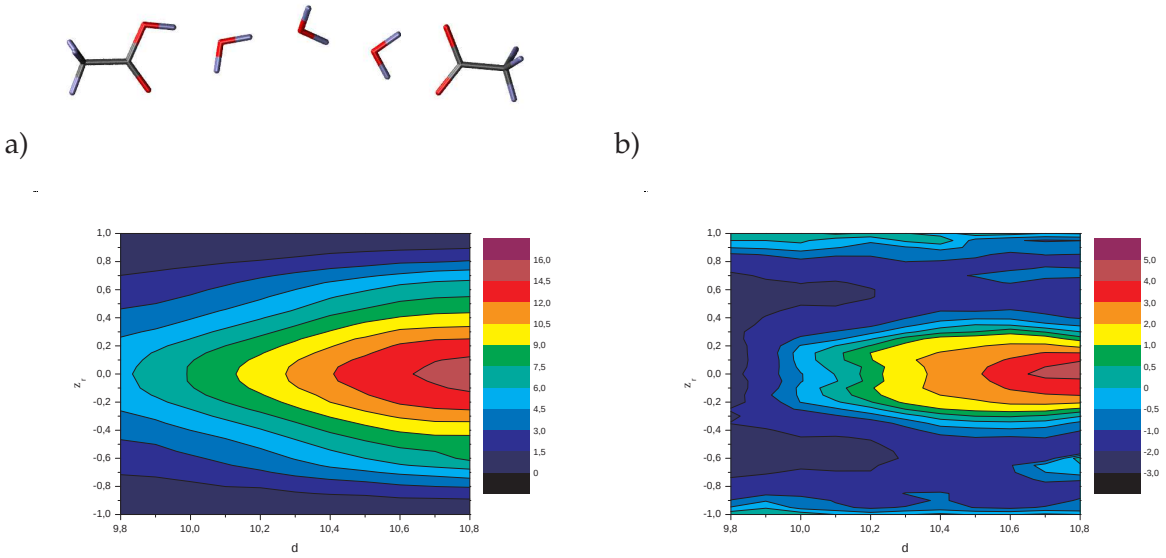


Figure A.3: Benchmark results for a three-water bridge between acetic acid and acetate. Energies (in kcal/mol) are relative to the energy of the reactand at ζ_R=-1.0 for the same value of d. a) SCC-DFTB energies along the reaction path for different donor-acceptor distances d, c) B3LYP/6-311G** - SCC-DFTB results.

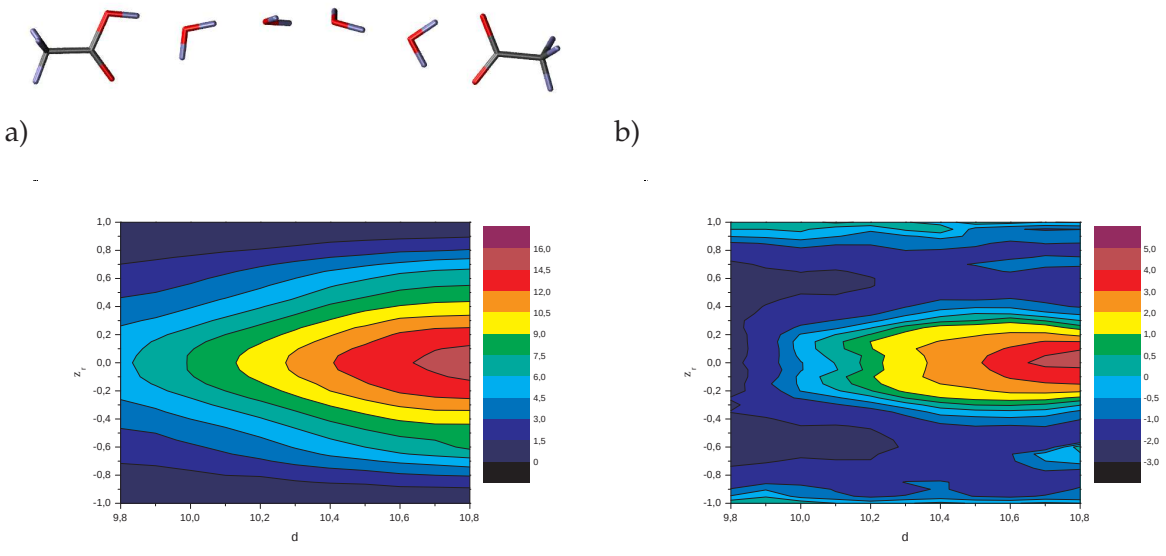


Figure A.4: Benchmark results for a four-water bridge between acetic acid and acetate. Energies (in kcal/mol) are relative to the energy of the reactand at ζ_R=-1.0 for the same value of d. a) SCC-DFTB energies along the reaction path for different donor-acceptor distances d, c) B3LYP/6-311G** - SCC-DFTB results.

Bibliography

[1] H. Gest. A misplaced chapter in the history of photosynthesis research; the second publication (1796) on plant processes by Dr Jan Ingen-Housz, MD, discoverer of photosynthesis – a bicentenniel resurrection howard gest. *Photosynth. Res.*, 53:65, 1997.

[2] R. J. Spreitzer and M. E. Salvucci. RUBISCO: Structure, regulatory interactions, and possibilities for a better enzyme. *Annu. Rev. Plant Biol.*, 53:449, 2002.

[3] D. Voet and Voet J. G. *Biochemistry*. Wiley, 2004.

[4] A.N. Glazer. Photochemical reaction centers: Structure, organization, and function. *Ann. Rev. Plant Physiol.*, 38:11, 1987.

[5] H. Komiya, T. O. Yeates, D. C. Rees, J. P. Allen, and G. Feher. Structure of the reaction center from rhodobacter sphaeroides R-26 and 2.4.1: 6. symmetry relations and sequence comparisons between different species. *Proc. Natl. Acad. Sci. USA*, 85:9012, 1988.

[6] J. Deisenhofer, O. Epp, K. Miki, and R. uber. Structure of the protein subunits in the photosynthetic reaction centre of Rhodopseudomonas viridis at 3å resolution. *Nature*, 318:618, 1985.

[7] J. P. Allen. My daily constitutional in Martinsried. *Photosynth. Res.*, 80:157, 2004.

[8] A. Fersht. *Structure and Mechanism in Protein Science. Guide to Enzyme Catalysis and Protein Folding*. W. H. Freeman, 1999.

[9] R. A. W. Frank, C. M. Titman, J. V. Pratap, B. F. Luisi, and R. N. Perham. A molecular switch and proton wire synchronize the active sites in thiamine enzymes. *Science*, 306:872–876, 2004.

[10] R. R. Birge. Nature of the primary photochemical events in rhodopsin and bacteriorhodopsin. *Biochim. Biophys. Acta*, 1016:293, 1990.

[11] M. Y. Okamura, M. L. Paddock, M. S. Graige, and G. Feher. Proton and electron transfer in bacterial reaction centers. *Biochim. Biophys. Acta*, 1458:148, 2000.

[12] M. Wikstrom. Cytochrome c oxidase: 25 years of the elusive proton pump. *Biochim. Biophys. Acta - Bioenergetics*, 1655:241–247, 2004.

[13] P. Brzezinski. Redox-driven membrane-bound proton pumps. *Trends in Biochem. Sci.*, 29:380, 2004.

[14] R. M. Stroud, L. J. W. Miercke, J. O'Connell, S. Khademi, J. K. Lee, J. Remis, W. Harries, Y. Robles, and D. Akhavan. Glycerol facilitator glpf and the associated aquaporin family of channels. *Curr. Opin. Struct. Biol.*, 13:424, 2003.

[15] M. Borgnia, S. Nielsen, A. Engel, and P. Agre. Cellular and molecular biology of the aquaporin water channels. *Annu. Rev. Biochem.*, 68:425, 1999.

[16] E. Tajkhorshid, P. Nollert, M. Ø. Jensen, L. J. W. Miercke, J. O'Connell, R. M. Stroud, and K. Schulten. Control of the selectivity of the aquaporin water channel family by global orientational tuning. *Science*, 296:525, 2002.

[17] N. Chakrabarti, E. Tajkhorshid, B. Roux, and R. Pomès. Molecular basis of proton blockage in aquaporins. *Structure*, 12:65, 2004.

[18] A. Burykin and A. Warshel. What really prevents proton transport through aquaporin? charge self-energy versus proton wire proposals. *Biophys. J.*, 85:3696, 2003.

[19] N. Chakrabarti, B. Roux, and R. Pomès. Structural determinants of proton blockage in aquaporins. *J. Mol. Biol.*, 343:493, 2004.

[20] R. Pomès and B. Roux. Structure and dynamics of a proton wire: a theoretical study of H^+ translocation along the single-file water chain in the gramicidin a channel. *Biophys. J.*, 71:19, 1996.

[21] R. Pomès and B. Roux. Molecular mechanism of H^+ conduction in the single-file water chain of the gramicidin channel. *Biophys. J.*, 82:2304, 2002.

[22] S. Braun-Sand, A. Burykin, Z. T. Chu, and A. Warshel. Realistic simulations of proton transport along the gramicidin channel: Demonstrating the importance of solvation effects. *J. Phys. Chem. B*, 109:583–592, 2005.

[23] K. Kreuer, S. J. Paddison, E. Spohr, and M. Schuster. Transport in proton conductors for fuel-cell applications: Simulations, elementary reactions, and phenomenology. *Chem. Rev.*, 104:4637–4678, 2004.

[24] A. Kohen and J. P. Klinman. Enzyme catalysis: Beyond classical paradigms. *Acc. Chem. Res.*, 31:397–404, 1998.

[25] Z. X. Liang and J. P. Klinman. Structural bases of hydrogen tunneling in enzymes: progress and puzzles. *Curr. Opin. Struct. Biol.*, 14:648–655, 2004.

[26] J. K. Hwang and A. Warshel. How important are quantum mechanical nuclear motions in enzyme catalysis. *J. Am. Chem. Soc.*, 118:11745–11751, 1996.

[27] J. Gao and D. G. Truhlar. Quantum mechanical methods for enzyme kinetics. *Annu. Rev. Phys. Chem.*, 53:467, 2002.

[28] Q. Cui and M. Karplus. Quantum mechanics/molecular mechanics studies of triosephosphate isomerase-catalyzed reactions: Effect of geometry and tunneling on proton-transfer rate constants. *J. Am. Chem. Soc.*, 124:3093, 2002.

[29] P. M. Kiefer and J. T. Hynes. Kinetic isotope effects for adiabatic proton transfer reactions in a polar environment. *J. Phys. Chem. A*, 107:9022–9039, 2003.

[30] S. Hammes-Schiffer. Impact of enzyme motion on activity. *Biochem.*, 41:13335–13343, 2002.

[31] A. Warshel. Computer simulations of enzyme catalysis: Methods, progress, and insights. *Annu. Rev. Biophys. Biomol. Struct.*, 32:425, 2002.

[32] S. J. Benkovic and S. Hammes-Schiffer. A perspective on enzyme catalysis. *Science*, 301:1196–1202, 2003.

[33] J. F. Nagle and H. J. Morowitz. Molecular mechanisms for proton transport in membranes. *Proc. Natl. Acad. Sci.*, 75:298–302, 1978.

[34] J. F. Nagle and M. Mille. Molecular models of proton pumps. *J. Chem. Phys.*, 74:1367, 1981.

[35] Y. Y. Sham, I. Muegge, and A. Warshel. Simulating proton translocations in proteins: probing proton transfer pathways in the rhodobacter sphaeroides reaction center. *Proteins: Struct., Funct., Genet.*, 36:484–500, 1999.

[36] M. Y. Okamura and G. Feher. Proton transfer in reaction centers from photosynthetic bacteria. *Annu. Rev. Biochemistry*, 61:861, 1992.

[37] P. Sebban, P. Maroti, and D. K. Hanson. Electron and proton transfer to the quinones in bacterial photosynthetic reaction centers: insight from combined approaches of molecular genetics and biophysics. *Biochimie*, 77:677, 1995.

[38] M. L. Paddock, G. Feher, and M. Y. Okamura. Proton transfer pathways and mechanism in bacterial reaction centers. *FEBS Lett.*, 555:45, 2003.

[39] J. Aqvist and A. Warshel. Computer simulation of the initial proton transfer step in human carbonic anhydrase. *J. Mol. Biol.*, 224:7–14, 1992.

[40] S. Braun-Sand, M. Strajbl, and A. Warshel. Studies of proton translocations in biological systems: simulating proton transport in carbonic anhydrase by evb-based models. *Biophys. J.*, 87:2221–2239, 2004.

[41] C. N. Schutz and A. Warshel. Analyzing free energy relationships for proton translocations in enzymes: carbonic anhydrase revisited. *J. Phys. Chem. B*, 108:2066–2075, 2004.

[42] D. Xu, D. Riccardi, N. Ghosh, M. Elstner, H. Guo, and Q. Cui. Concerted proton transfers in carbonic anhydrase revealed by QM/MM simulations. *to be submitted*.

[43] S. Toba, G. Colombo, and K. M. Jr. Merz. Solvent dynamics and mechanism of proton transfer in human carbonic anhydrase ii. *J. Am. Chem. Soc.*, 121:2290–2302, 1999.

[44] D. Lu and G. A. Voth. Molecular dynamics simulations of human carbonic anhydrase ii: insight into experimental results and the role of solvation. *Proteins: Struct., Funct., and Genet.*, 33:119–134, 1998.

[45] R. Pomès and B. Roux. Free energy profiles for H^+ conduction along hydrogen-bonded chains of water molecules. *Biophys. J.*, 75:33–40, 1996.

[46] M. F. Schumaker, R. Pomès, and B. Roux. A combined molecular dynamics and diffusion model of single proton conduction through gramicidin. *Biophys. J.*, 79:2840, 2000.

[47] A. Bondar, S. Fischer, J. C. Smith, M. Elstner, and S. Suhai. Key role of electrostatic interactions in bacteriorhodopsin proton transfer. *J. Am. Chem. Soc.*, 126:14668, 2004.

[48] A. Bondar, M. Elstner, S. Suhai, J. C. Smith, and S. Fischer. Mechanism of primary proton transfer in bacteriorhodopsin. *Structure*, 12:1281, 2004.

[49] Y. Wu and G. A. Voth. A computer simulation study of the hydrated proton in a synthetic proton channel. *Biophys. J.*, 85:864, 2003.

[50] D. M. Popovic and A. A. Stuchebrukhov. Electrostatic study of the proton pumping mechanism in bovine heart cytochrome c oxidase. *J. Am. Chem. Soc.*, 126:1858–1871, 2004.

[51] R. I. Cukier. A molecular dynamics study of water chain formation in the proton-conducting k channel of cytochrome c oxidase. *Biochim. Biophys. Acta.*, 1706:134–146, 2005.

[52] E. Olkhova, M. C. Huter, M. A. Lill, V. Helms, and H. Michel. Dynamic water networks in cytochrome c oxidase from paracoccus denitrificans investigated by molecular dynamics simulations. *Biophys. J.*, 86:1873–1889, 2004.

[53] M. Wikstrom, M. I. Verkhovsky, and G. Hummer. Water gated mechanism of proton translocation by cytochrome c oxidase. *Biochim. Biophys. Acata*, 1604:61–65, 2003.

[54] G. H. Johannessón and H. Jónsson. Optimization of hyperplanar transition states. *J. Chem. Phys.*, 115:9644, 2001.

[55] Stefan Fischer and Martin Karplus. Conjugate Peak refinement: an algorithm for finding reaction paths and accurate transition states in systems with many degrees of freedom. *Chem. Phys. Lett.*, 194(3):252, 1992.

[56] M. Klähn, S. Braun-Sand, E. Rosta, and A. Warshel. On possible pitfalls in ab initio quantum mechanics/molecular mechanics minimization approaches for studies of enzymatic reactions. *J. Phys. Chem. B*, 109:15645, 2005.

[57] J. Baschnagel, K. Binder, P. Doruker, A. A. Gusev, O. Hahn, K. Kremer, W. L. Mattice, F. Müller-Plathe, M. Murat, W. Paul, S. Santos, U. W. Suter, and V. Tries. Bridging the gap between atomistic and coarse-grained models of polymers: Status and perspectives. *Adv. Polymer Sci.*, 152:41, 2000.

[58] A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher III, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, and M. Karplus. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. of Phys. Chem. B*, 102:3586, 1998.

[59] M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, Th. Frauenheim, S. Suhai, and G. Seifert. Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties. *Phys. Rev. B*, 58:7260, 1998.

[60] Q. Cui, M. Elstner, E. Kaxiras, Th. Frauenheim, and M. Karplus. A QM/MM implementation of the self-consistent charge density functional tight binding (SCC-DFTB) method. *J. Phys. Chem. B*, 105:569, 2001.

[61] B. R. Brooks, R. E. Bruccoleri, B. E. Olafson, D. J. States, S. Swaminathan, and M. Karplus. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.*, 4:187, 1983.

[62] M. Elstner, Th. Frauenheim, and S. Suhai. An approximate DFT method for QM/MM simulations of biological structures and processes. *THEOCHEM*, 632:29, 2003.

[63] M. Elstner, Q. Cui, P. Munih, E. Kaxiras, Th. Frauenheim, and M. Karplus. Modeling zinc in biomolecules with the self consistent charge-density functional tight binding (SCC-DFTB) method: Applications to structural and energetic analysis. *J. Comput. Chem.*, 24:565, 2003.

[64] X. Zhang, D. H. T. Harrison, and Q. Cui. Functional specificities of methylglyoxal synthase and triosephosphate isomerase: A combined QM/MM analysis. *J. Am. Chem. Soc.*, 124:14871, 2002.

[65] M. S. Formaneck, G. Li, X. Zhang, and Q. Cui. Calculating accurate redox potentials in enzymes with a combined QM/MM free energy perturbation approach. *J. Theor. Comput. Chem.*, 1:53, 2002.

[66] Q. Cui, M. Elstner, and M. Karplus. A theoretical analysis of the proton and hydride transfer in liver alcohol dehydrogenase (LADH). *J. Phys. Chem. B*, 106:2721, 2002.

[67] G. Li and Q. Cui. What is so special about Arg 55 in the catalysis of cyclophilin A? insights from hybrid QM/MM simulations. *J. Am. Chem. Soc.*, 125:15028, 2003.

[68] G. Li, X. Zhang, and Q. Cui. Free energy perturbation calculations with combined QM/MM potentials complications, simplifications, and applications to redox potential calculations. *J. Phys. Chem. B*, 107:8643, 2003.

[69] G. Li and Q. Cui. pK_a calculations with QM/MM free energy perturbations. *J. Phys. Chem. B*, page 14521, 2003.

[70] E. Granseth, H. G. Von, and A. Elofsson. A study of the membrane-water interface region of membrane proteins. *J. Mol. Biol.*, 346:377, 2005.

[71] T. Helgaker, P. Jørgensen, and J. Olsen. *Molecular Electronic-Structure Theory*. John Wiley & Sons, 2000.

[72] P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Phys Rev. B*, 136:864–871, 1964.

[73] R. O. Jones and O. Gunnarsson. The density functional formalism, its applications and prospects. *Rev. Mod. Phys.*, 61:689, 1989.

[74] W. Kohn and Sham L. J. Self-consistent equations including exchange and correlation effects. *Phys. Rev.*, 140:1133, 1965.

[75] K. Capelle. A bird's-eye view of density-functional theory. *Los Alamos National Laboratory, Preprint Archive, Condensed Matter*, pages 1, arXiv:cond–mat/0211443, 2005.

[76] A. D. Becke. Density-functional thermochemistry. III. the role of exact exchange. *J. Chem. Phys.*, 98:5648, 1993.

[77] R. A. Friesner. Ab initio quantum chemistry: Methodology and applications. *Proc. Natl. Acad. Sci.*, 102:6648, 2005.

[78] M. Ø. Jensen, U. Röthlisberger, and C. Rovira. Hydroxide and proton migration in aquaporins. *Biophys. J.*, 89:1744–1759, 2005.

[79] Th. Frauenheim, F. Weich, Th. Köhler, S. Uhlmann, D. Porezag, and G. Seifert. Density-functional based construction of transferable non-orthogonal tight-binding potentials for Si and SiH. *Phys. Rev. B*, 52:11492, 1995.

[80] M. Sternberg. *The Atomic Structure of Diamond Surfaces and Interfaces*. PhD thesis, Universität Paderborn, Paderborn, Germany, 2001.

[81] Th. Frauenheim, G. Seifert, M. Elstner, Th. Niehaus, and C. Köhler. Atomistic simulations of complex materials: ground-state and excited-state properties. *J. Phys. Cond. Matt.*, 14:3015, 2002.

[82] G. Seifert, H. Eschrig, and W. Bieger. Eine approximative Variante des LCAO-Xα- Verfahrens. *Z. Phys. Chemie*, 267:529, 1986.

[83] W. M. C. Foulkes and R. Haydock. Tight-binding models and density functional theory. *Phys. Rev. B*, 39:12520, 1989.

[84] J. Elsner. *Surfaces and Extended Defects in Wurtzite GaN*. PhD thesis, Universität Paderborn, Paderborn, Germany, 1998.

[85] N. L. Allinger. Conformational analysis. 130. MM2. a hydrocarbon force field utilizing v1 and v2 torsional terms. *Journal of the American Chemical Society*, 99:8127, 1977.

[86] N. L. Allinger, Y. H. Yuh, and J. H. Lii. Molecular mechanics. the MM3 force field for hydrocarbons. 1. *J. Am. Chem. Soc.*, 111:8551, 1989.

[87] A. D. MacKerell. Empirical force fields: Overview and parameter optimization. In *43th Sanibel Symposium*, 2003.

[88] D. A. Pearlman, D. A. Case, J. W. Caldwell, W. S. Ross, T. E. I. Cheatham, S. DeBolt, D. Ferguson, G. Seibel, and P. Kollman. "AMBER", a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to stimulate the structural and energetic properties of molecules. *Comput. Phys. Commun.*, 91:1, 1995.

[89] J. W. Ponder and D. A. Case. Force fields for protein simulations. *Adv. Protein Chem.*, 66:27, 2003.

[90] A. D. J. MacKerell, M. Feig, and C. L. Brooks III. Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.*, 25:1400, 2004.

[91] F. Jensen. *Introduction to computational chemistry*. John Wiley & Sons, 1999.

[92] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, 79:926, 1983.

[93] C. I. Bayly, P. Cieplak, W. Cornell, and P. A. Kollman. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J. Phys. Chem.*, 97:10269, 1993.

[94] L. E. Chirlian and M. M. Francl. Atomic charges derived from electrostatic potentials: a detailed study. *J. Comput. Chem.*, 8:894, 1987.

[95] C. M. Breneman and K. B. Wiberg. Determining atom-centered monopoles from molecular electrostatic potentials. the need for high sampling density in formamide conformational analysis. *J. Comput. Chem.*, 11:361, 1990.

[96] W. Yang and T.-S. Lee. A density-matrix divide-and-conquer approach for electronic structure calculations of large molecules. *J. Chem. Phys.*, 103:5674, 1995.

[97] H. Liu, M. Elstner, E. Kaxiras, Th. Frauenheim, J. Hermans, and W. Yang. Quantum mechanics simulation of protein dynamics on long timescale. *Proteins*, 44:484, 2001.

[98] G. E. Moore. Cramming more components onto integrated circuits. *Electronics*, 38, 1965.

[99] J. J. P. Stewart. Optimization of parameters for semiempirical methods. I. Method. *J. Comput. Chem.*, 10:209, 1989.

[100] M. J. S. Dewar, E. G. Zoebisch, E. F. Healy, and J. J. P. Stewart. Development and use of quantum mechanical molecular models. AM1: a new general purpose quantum mechanical molecular model. *J. Am.Chem. Soc.*, 107:3902, 1985.

[101] T. A. Weber and F. H. Stillinger. Dynamical study of the $H_5O_2^+ + H_3O_2^-$ neutralization reaction using the polarization model. *J. Chem. Phys.*, 77:4150, 1982.

[102] M. J. Field, P. A. Bash, and M. Karplus. A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations. *J. Comput. Chem.*, 11:700, 1990.

[103] J. Aaqvist and A. Warshel. Simulation of enzyme reactions using valence bond force fields and other hybrid quantum/classical approaches. *Chem. Rev.*, 93:2523, 1993.

[104] J. Gao. Hybrid quantum mechanical/molecular mechanical simulations: An alternative avenue to solvent effects in organic chemistry. *Acc. Chem. Res.*, 29:298, 1996.

[105] Q. Cui and M. Karplus. Catalysis and specificity in enzymes: a study of triosephosphate isomerase and comparison with methyl glyoxal synthase. *Adv. Prot. Chem.*, 66:315, 2003.

[106] C. Lennartz, A. Schaefer, F. Terstegen, and W. Thiel. Enzymatic reactions of triosephosphate isomerase: A theoretical calibration study. *J. Phys. Chem. B*, 106:1758, 2002.

[107] M. Svensson, S. Humbel, R. D. J. Froese, T. Matsubara, S. Sieber, and K. Morokuma. ONIOM: A multilayered integrated MO + MM method for geometry optimizations and single point energy predictions. a test for diels-alder reactions and $Pt(P(t\text{-}Bu)_3)_2 + H_2$ oxidative addition. *J. Phys. Chem.*, 100:19357, 1996.

[108] D. Riccardi, G. Li, and Q. Cui. Importance of van der waals interactions in QM/MM simulations. *J. Phys. Chem. B*, 108:6467, 2004.

[109] D. Bakowies and W. Thiel. Hybrid models for combined quantum mechanical and molecular mechanical approaches. *J. Phys. Chem.*, 100:10580, 1996.

[110] J. Sauer and M. Sierka. Combining quantum mechanics and interatomic potential functions in ab initio studies of extended systems. *J. Comput. Chem.*, 21:1470–1493, 2000.

[111] M. Garcia-Viloca, D. G. Truhlar, and J. Gao. Importance of substrate and cofactor polarization in the active site of dihydrofolate reductase. *J. Mol. Biol.*, 327:549, 2003.

[112] G. Lamoureux and B. Roux. Modeling induced polarization with classical drude oscillators: Theory and molecular dynamics simulation algorithm. *J. Chem. Phys.*, 119:3025–3039, 2003.

[113] A. K. Rappe and W. A. Goddard III. Charge equilibration for molecular dynamics simulations. *J. Phys. Chem.*, 95:3358–3363, 1991.

[114] Q. Cui, M. Elstner, E. Kaxiras, Th. Frauenheim, and M. Karplus. A QM/MM implementation of the self-consistent charge density functional tight binding (SCC-DFTB) method. *J. Phys. Chem. B*, 105:569, 2001.

[115] D. Das, K. Eurenius, E. Billings, P. Sherwood, D. Chatfield, M. Hodoscek, and B. Brooks. Optimization of quantum mechanical molecular mechanical partitioning schemes: Gaussian delocalization of molecular mechanical charges and the double link atom method. *J.Chem.Phys.*, 117:10534, 2002.

[116] P. H. König, M. Hoffmann, Th. Frauenheim, and Q. Cui. A critical evaluation of different QM/MM frontier treatments using SCC-DFTB as the QM method. *J. Phys. Chem. B*, 109:9082–9095, 2005.

[117] P. J. Steinbach and Brooks B. R. New spherical-cutoff methods for long-range forces in macromolecular simulation. *J. Comput. Chem.*, 15:667 – 683, 1994.

[118] R. H. Stote and M. Karplus. Zinc binding in proteins and solution: A simple but accurate nonbonded representation. *Proteins*, 23:12, 1995.

[119] R. H. Stote, D. J. States, and M. Karplus. On the treatment of electrostatic interactions in biomolecular simulation. *J. Chim. Phys.*, 88:2419–33, 1991.

[120] D. Sengupta, R. N. Behera, J. C. Smith, and G. M. Ullmann. The a helix dipole: Screened out? *Structure*, 13:849, 2005.

[121] K. Nam, J. Gao, and D. M. York. An efficient linear-scaling ewald method for long-range electrostatic interactions in combined QM/MM calculations. *J. Chem. Theo. Comput.*, 1:2, 2005.

[122] P. Schaefer, D. Riccardi, and Q. Cui. Reliable treatment of electrostatics in combined QM/MM simulation of macromolecules. *J. Chem. Phys.*, 123:014905, 2005.

[123] D. D. Riccardi, P. Schaefer, and Q. Cui. pK_a calculations in solution and proteins with QM/MM free energy perturbation simulations: A quantitative test of QM/MM protocols. *J. Phys. Chem. B*, 109:17715, 2005.

[124] P. H. König, N. Ghosh, M. Hoffmann, M. Elstner, E. Tajkhorshid, Th. Frauenheim, and Q. Cui. Towards theoretical analysis of proton transfer kinetics in biomolecular pumps. in print, DOI: 10.1021/jp0442347, 2005.

[125] A. Shurki and A. Warshel. Structure/function correlations of proteins using MM, QM/MM, and related approaches: methods, concepts, pitfalls, and current progress. *Adv. Prot. Chem.*, 66:249, 2003.

[126] T. Simonson, G. Archontis, and M. Karplus. Continuum treatment of long-range interactions in free energy calculations. application to protein–ligand binding. *J. Phys. Chem. B*, 101:8349, 1997.

[127] C. L. Brooks and M. Karplus. Solvent effects on protein motion and protein effects on solvent motion. dynamics of the active site region of lysozyme. *J. Mol. Biol.*, 208:159, 1989.

[128] C. Sagui and T. A. Darden. Molecular dynamics simulations of biomolecules: Long-range electrostatic effects. *Annu. Rev. Biophys. Biomol. Struct.*, 28:155–179, 1999.

[129] S. Kuwajima and A. Warshel. The extended Ewald method: a general treatment of long-range electrostatic interactions in microscopic simulations. *J. Chem. Phys.*, 89:3751, 1988.

[130] W. Im, S. Berneche, and B. Roux. Generalized solvent boundary potential for computer simulations. *J. Chem. Phys.*, 114:2924, 2001.

[131] D. Frenkel and B. Smit. *Understanding Molecular Simulation : From Algorithms to Applications*. Academic Press, 2001.

[132] A. Leach. *Molecular Modelling: Principles and Applications*. Prentice Hall, 2001.

[133] B. R. Gelin and M. Karplus. Sidechain torsional potentials and motion of amino acids in proteins. bovine pancreatic trypsin inhibitor. *Proc. Natl. Acad. Sci.*, 72:2002, 1975.

[134] S. Fischer, R. L. Dunbrack, and M. Karplus. Cis- trans imide isomerization of the proline dipeptide. *J. Am. Chem. Soc.*, 116:11931, 1994.

[135] R. Czerminski and R. Elber. Self-avoiding walk between two fixed points as a tool to calculate reaction paths in large molecular systems. *Int. J. Quantum Chem.*, 24:167, 1990.

[136] P. Maragakis, S. A. Andreev, Y. Brumer, D. R. Reichman, and E. Kaxiras. Adaptive nudged elastic band approach for transition state calculation. *J. Chem. Phys.*, 117:4651, 2002.

[137] D. N. Silverman. Marcus rate theory applied to enzymatic proton transfer. *Biochim. Biophys. Acta.*, 1458:88–103, 2000.

[138] D. Riccardi and Cui Q. in preparation.

[139] A. R. Leach. *Molecular Modelling: Priciples and application*. Prentice Hall, 2001.

[140] J. P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen. Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of N-alkanes. *J. Comput. Phys.*, 23:327, 1977.

[141] L. S. D. Caves, J. D. Evanseck, and M. Karplus. Locally accessible conformations of proteins: Multiple molecular dynamics simulations of crambin. *Biochim. Biophys. Acta*, 7:649, 1998.

[142] H. C. Andersen. Molecular dynamics at constant pressure and/or temperature. *J. Chem. Phys.*, 72:2384, 1980.

[143] Nosé. A molecular dynamics method for simulation in the canonical ensemble. *Mol. Phys.*, 52:255, 1984.

[144] W. G. Hoover. A molecular dynamics method for simulation in the canonical ensemble. *Phys. Rev. A*, 31:1695, 1985.

[145] H. Grubmüller. Predicting slow structural transitions in macromolecular systems: conformational flooding. *Phys. Rev. E*, 52:2893, 1995.

[146] A. F. Voter. Hyperdynamics: accelerated molecular dynamics of infrequent events. *Phys. Rev. Lett.*, 78:3908, 1997.

[147] C. L. Brooks, M. Karplus, and B. M. Pettitt. *Proteins: A Theoretical Perspective of Dynamics, Structure, and Thermodynamics*. Wiley, 1990.

[148] D. Chandler. *Introduction to Modern Statistical Mechanics*. Oxford University Press, 1987.

[149] S. Kumar, D. Bouzida, R. H. Swendsen, P. A. Kollman, and J. M. Rosenberg. The weighted histogram analysis method for free energy calculations on biomolecules. *J. Comput. Chem.*, 13:1011, 1992.

[150] Y. Zhang, T. Lee, and W. Yang. A pseudobond approach to combining quantum mechanical and molecular mechanical methods. *J. Chem. Phys.*, 110:46, 1999.

[151] I. Antes and W. Thiel. Adjusted connection atoms for combined quantum mechanical and molecular mechanical methods. *J. Phys. Chem. A*, 103:9290, 1999.

[152] G. G. Ferenczy, J. L. Rivail, P. R. Surjan, and G. Naray-Szabo. NDDO fragment self-consistent field approximation for large electronic systems. *J. Comput. Chem.*, 13:830, 1992.

[153] V. Thery, D. Rinaldi, J. L. Rivail, B. Maigret, and G. G. Ferenczy. Quantum-mechanical computations on very large molecular systems: the local self-consistent field method. *J. Comput. Chem.*, 15:269, 1994.

[154] D. M. Philipp and R. A. Friesner. Mixed ab initio QM/MM modeling using frozen orbitals and tests with alanine dipeptide and tetrapeptide. *J. Comput. Chem.*, 20:1468, 1999.

[155] R. B. Murphy, D. M. Philipp, and R. A. Friesner. A mixed quantum mechanics/molecular mechanics (QM/MM) method for large-scale modeling of chemistry in protein environments. *J. Comput. Chem.*, 21:1442, 2000.

[156] J. Gao, P. Amara, C. Alhambra, and M. J. Field. A generalized hybrid orbital (GHO) method for the treatment of boundary atoms in combined QM/MM calculations. *J. Phys. Chem. A*, 102:4714, 1998.

[157] X. Assfeld and J. Rivail. Quantum chemical computations on parts of large molecules: the ab initio local self consistent field method. *Chem. Phys. Lett.*, 263:100, 1996.

[158] N. Ferre, X. Assfeld, and J. Rivail. Specific force field parameters determination for the hybrid ab initio QM/MM LSCF method. *J. Comput. Chem.*, 23:610, 2002.

[159] X. Assfeld, N. Ferre, and J. Rivail. Electrostatic interactions in peptides. polarisation effects due to an α-helix. *Theo. Chem. Acc.*, 111:328, 2004.

[160] N. Reuter, A. Dejaegere, B. Maigret, and M. Karplus. Frontier bonds in QM/MM methods: A comparison of different approaches. *J. Phys. Chem. A*, 104:1720, 2000.

[161] M. J. Harrison, N. A. Burton, and I. H. Hillier. Catalytic mechanism of the enzyme papain: Predictions with a hybrid quantum mechanical/molecular mechanical potential. *J. Am. Chem. Soc.*, 119:12285, 1997.

[162] V. V. Vasilyev. Tetrahedral intermediate formation in the acylation step of acetylcholinesterases. a combined quantum chemical and molecular mechanical model. *THEOCHEM*, 110:129, 1994.

[163] N. Ferre and M. Olivucci. The amide bond: pitfalls and drawbacks of the link atom scheme. *THEOCHEM*, 632:71, 2003.

[164] B. Waszkowycz, I. H. Hillier, N. Gensmantel, and D. W. Payling. A combined quantum mechanical/molecular mechanical model of the potential energy surface of ester hydrolysis by the enzyme phospholipase A2. *Perkin Trans. 2*, page 225, 1991.

[165] B. Waszkowycz, I. H. Hillier, N. Gensmantel, and D. W. Payling. A quantum mechanical/molecular mechanical model of inhibition of the enzyme phospholipase A2. *Perkin Trans. 2*, page 1819, 1991.

[166] B. Waszkowycz, I. H. Hillier, N. Gensmantel, and D. W. Payling. Combined quantum mechanical-molecular mechanical study of catalysis by the enzyme phospholipase A2: an investigation of the potential energy surface for amide hydrolysis. *Perkin 2*, page 2025, 1991.

[167] U. C. Singh and P. A. Kollmann. A combined ab initio quantum mechanical and molecular mechanical method for carrying out simulations on complex molecular systems: applications to the $CH_3Cl + Cl^-$ exchange reaction and gas phase protonation of polyethers. *J. Comput. Chem.*, 7:718, 1986.

[168] A. H. de Vries, P. Sherwood, S. J. Collins, A. M. Rigby, M. Rigutto, and G. J. Kramer. Zeolite structure and reactivity by combined quantum-chemical-classical calculations. *J. Phys. Chem. B*, 103:6133, 1999.

[169] P. Sherwood, A. H. de Vries, S. J. Collins, S. P. Greatbanks, N. A. Burton, M. A. Vincent, and I. H. Hillier. Computer simulation of zeolite structure and reactivity using embedded cluster methods. *Faraday*, 106:79, 1997.

[170] P. Amara and M. J. Field. Evaluation of an ab initio quantum

mechanical/molecular mechanical hybrid-potential link-atom method. *Theor. Chem. Acc.*, 109:43, 2003.

[171] J. Pu, J. Gao, and D. G. Truhlar. Combining self-consistent-charge density-functional tight-binding (SCC-DFTB) with molecular mechanics by the generalized hybrid orbital (GHO) method. *J. Phys. Chem. A*, 108:5454, 2004.

[172] Q. Cui and M. Karplus. Triosephosphate isomerase: A theoretical comparison of alternative pathways. *J. Am. Chem. Soc.*, 123:2284, 2001.

[173] Q. Cui and M. Karplus. Quantum mechanical/molecular mechanical studies of the triosephosphate isomerase-catalyzed reaction: Verification of methodology and analysis of reaction mechanisms. *J. of Phys. Chem. B*, 106:1768, 2002.

[174] J. J. P. Stewart. Semiempirical molecular orbital methods. *Rev. Comput. Chem.*, 1:45, 1990.

[175] R. C. Davenport, P. A. Bash, B. A. Seaton, M. Karplus, G. A. Petsko, and D. Ringe. Structure of the triosephosphate isomerase-phosphoglycolohydroxamate complex: an analog of the intermediate on the reaction pathway. *Biochemistry*, 30:5821, 1991.

[176] E. Lolis, T. Alber, R. C. Davenport, D. Rose, F. C. Hartman, and G. A. Petsko. Structure of yeast triosephosphate isomerase at 1.9 å resolution. *Biochemistry*, 29:6609, 1997.

[177] T. K. Harris, C. Abeygunawardana, and A. S. Mildvan. Nmr studies of the role of hydrogen bonding in the mechanism of triosephosphate isomerase. *Biochemistry*, 36:14661, 1997.

[178] T. K. Harris, R. N. Cole, F. I. Comer, and A. S. Mildvan. Proton transfer in the mechanism of triosephosphate isomerase. *Biochemistry*, 37:16828, 1998.

[179] E. B. Nickbarg, R. C. Davenport, G. A. Petsko, and J. R. Knowles. Triosephosphate isomerase: removal of a putatively electrophilic histidine residue results in a subtle change in catalytic mechanism. *Biochemistry*, 27:5948, 1977.

[180] J. R. Knowles and W. J. Albery. Perfection in enzyme catalysis: the energetics of triosephosphate isomerase. *Acc. Chem. Res.*, 10:105–11, 1977.

[181] W. J. Albery and J. R. Knowles. Free-energy profile of the reaction catalyzed by triosephosphate isomerase. *Biochemistry*, 15:5627, 1976.

[182] W. J. Albery and J. R. Knowles. Deuterium and tritium exchange in enzyme kinetics. *Biochemistry*, 15:5588, 1976.

[183] S. G. Maister, C. P. Pett, W. J. Albery, and J. R. Knowles. Energetics of triosephosphate isomerase: the appearance of solvent tritium in substrate dihydroxyacetone phosphate and in product. *Biochemistry*, 15:5607, 1976.

[184] J. M. Herlihy, S. G. Maister, W. J. Albery, and J. R. Knowles. Energetics of triosephosphate isomerase: the fate of the 1(R)-^{3}h label of tritiated dihydroxyacetone phsophate in the isomerase reaction. *Biochemistry*, 15:5601, 2000.

[185] M. Perakyla and T. A. Pakkanen. Ab initio models for receptor-ligand interactions in proteins. 4. model assembly study of the catalytic mechanism of triosephosphate isomerase. *Proteins*, 25:225, 1996.

[186] G. Alagona, P. Desmeules, C. Ghio, and P. A. Kollman. Quantum mechanical and molecular mechanical studies on a model for the dihydroxyacetone phosphate-glyceraldehyde phosphate isomerization catalyzed by triose phosphate isomerase (TIM). *J. Am. Chem. Soc.*, 106:3623, 1984.

[187] G. Alagona, C. Ghio, and P. A. Kollman. Simple model for the effect of the Glu165 -> Asp165 mutation on the rate of catalysis in triose phosphate isomerase. *J. Mol. Biol.*, 191:23, 1986.

[188] G. Alagona, C. Ghio, and P. A. Kollman. Do enzymes stabilize transition states by electrostatic interactions or pK_a balance: The case of triose phosphate isomerase (TIM)? *J. Am. Chem. Soc.*, 117:9855, 1995.

[189] P. A. Bash, M. J. Field, R. C. Davenport, G. A. Petsko, D. Ringe, and M. Karplus. Computer simulation and analysis of the reaction pathway of triosephosphate isomerase. *Biochemistry*, 30:5826–32, 1991.

[190] I. Feierberg, V. Luzhkov, and J. Aqvist. Tim. *J. Biol. Chem.*, 275:22657, 2000.

[191] D. Saadat and D. H. T. Harrison. Mirroring perfection: The structure of methylglyoxal synthase complexed with the competitive inhibitor 2-phosphoglycolate. *Biochemistry*, 39:2950, 2000.

[192] G. T. Marks, T. K. Harris, M. A. Massiah, A. S. Mildvan, and D. H. T. Harrison. Mechanistic implications of methylglyoxal synthase complexed with phosphoglycolohydroxamic acid as observed by x-ray crystallography and NMR spectroscopy. *Biochemistry*, 40:6805, 2001.

[193] H. Liu, Z. Lu, G. A. Cisneros, and W. Yang. Parallel iterative reaction path optimization in ab initio quantum mechanical/molecular mechanical modeling of enzyme reactions. *J. Chem. Phys.*, 121:697, 2004.

[194] Y. Zhang, H. Liu, and W. Yang. Free energy calculation on enzyme reactions with an efficient iterative procedure to determine minimum energy paths on a combined ab initio QM/MM potential energy surface. *J. Chem. Phys.*, 112:3483–3492, 2000.

[195] Q. Cui and M. Karplus. Molecular properties from combined QM/MM methods. i. analytical second derivative and vibrational calculations. *J. Chem. Phys.*, 112:1133, 2000.

[196] Q. Cui and M. Karplus. Molecular properties from combined QM/MM methods. 2. Chemical shifts in large molecules. *J. Phys. Chem. B*, 104:3721, 2000.

[197] M. Nonella, G. Mathias, M. Eichinger, and P. Tavan. Structures and vibrational frequencies of the quinones in Rb. sphaeroides derived by a combined density functional/molecular mechanics approach. *J. Phys. Chem. B*, 107:316, 2003.

[198] M. Klähn, G. Mathias, C. Koetting, M. Nonella, J. Schlitter, K. Gerwert, and P. Tavan. IR spectra of phosphate ions in aqueous solution: Predictions of a DFT/MM approach compared with observations. *J. Phys. Chem. A*, 108:6186, 2004.

[199] B. A. Gregersen and D. M. York. Variational electrostatic projection (VEP) methods for efficient modeling of the macromolecular electrostatic and solvation environment in activated dynamics simulations. *J. Phys. Chem. B*, 109:536, 2005.

[200] P. A. Bash, L. L. Ho, A. D. MacKerell, D. Levine, and P. Hallstrom. Progress toward chemical accuracy in the computer simulation of condensed phase reactions. *Proc. Natl. Acad. Sci.*, 93:3698, 1996.

[201] A. M. Smondyrev and G. A. Voth. Molecular dynamics simulation of proton transport through the influenza a virus m2 channel. *Biophys. J.*, 83:1987–1996, 2002.

[202] P. G. Bolhuis, D. Chandler, C. Dellago, and P. L. Geissler. Transition path sampling: throwing ropes over rough mountain passes, in the dark. *Annu. Rev. Phys. Chem.*, 53:291, 2002.

[203] Y. R. Mo and J. Gao. Ab initio QM/MM simulations with a molecular orbital-valence bond (MOVB) method: applicatiob to an SN_2 reaction in water. *J. Comput. Chem.*, 21:1458–1469, 2000.

[204] E. Neria and M. Karplus. Molecular dynamics of an enzyme reaction: pro-

ton transfer in tim. *Chem. Phys. Lett.*, 267:23, 1997.

[205] K. Nam, X. Prat-Resina, M. Garcia-Viloca, L. S. Devi-Kesavan, and J. Gao. Dynamics of an enzymatic substitution reaction in haloalkane dehalogenase. *J. Am. Chem. Soc.*, 126:1369, 2004.

[206] D. Marx, M. E. Tuckerman, J. Hutter, and M. Parrinello. The nature of the hydrated excess proton in water. *Nature*, 397:601, 1999.

[207] M. Eigen. Proton transfer, acid-base catalysis, and enzymic hydrolysis. i. elementary processes. *Angew. Chem.*, 75:489–508, 1963.

[208] G. Zundel and H. Metzger. Energy bands of tunneling excess protons in liquid acids. IR spectroscopic study of the nature of $H_5O_2^+$ groups. *Z. Phys. Chemie*, 58:225, 1968.

[209] R. Pomès and B. Roux. Free energy profiles for H^+ conduction along hydrogen-bonded chains of water molecules. *Biophys. J.*, 75:33, 1998.

[210] R. Elber. Calculation of the potential of mean force using molecular dynamics with linear constraints. *J. Chem. Phys.*, 93:4312–4321, 1990.

[211] L. Backman. Binding of human carbonic anhydrase to human hemoglobin. *Eur. J. Biochem.*, 120:257, 1981.

[212] D. N. Silverman. Proton transfer in carbonic anhydrase measured by equilibrium isotope exchange. *Methods in Enzymology*, 249:479–503, 1995.

[213] D. W. Christianson and C. A. Fierke. Carbonic anhydrase: Evolution of the zinc binding site by nature and by design. *Acc. Chem. Res.*, 29:331–339, 1996.

[214] C. Tu, D. N. Silverman, C. Forsman, B. H. Jonsson, and S. Lindskog. Role of histidine 64 in the catalytic mechanism of human carbonic anhydrase II studied with a site-specific mutant. *Biochemistry*, 28:7913–18, 1989.

[215] A. E. Eriksson, T. A. Jones, and A. Liljas. Refined structure of human carbonic anhydrase ii at 2.0 a resolution. *Proteins*, 4:274–82., 1988.

[216] Z. Fisher, J. A. Hernandez Prada, C. Tu, D. Duda, C. Yoshioka, An. H., L. Govindasamy, D. N. Silverman, and R. McKenna. Structural and kinetic characterization of active-site histidine as a proton shuttle in catalysis by human carbonic anhydrase ii. *Biochemistry*, 44:1097, 2005.

[217] Q. Cui and M. Karplus. Is a "proton wire´´ concerted or stepwise? a model study of proton transfer in carbonic anhydrase. *J. Phys. Chem. B*, 107:1071, 2003.

[218] D. Borgis and J. T. Hynes. Dynamic theory of proton tunneling transfer rates in solution - general formulation. *Chem. Phys.*, 170:315–346, 1993.

[219] W. Kuhlbrandt. Structure and function of bacterial light-harvesting complexes. *Structure*, 3:521, 1995.

[220] A. Hiraishi. Transfer of the bacteriochlorophyll b-containing phototrophic bacteria rhodopseudomonas viridis and rhodopseudomonas sulfoviridis to the genus blastochloris gen. nov. *Int. J. System. Bacteriol.*, 47:217, 1997.

[221] R. J. Debus, G. Feher, and M. Y. Okamura. Iron-depleted reaction centers from rhodopseudomonas sphaeroides R-26.1: characterization and reconstitution with iron(2+), manganese(2+), cobalt(2+), nickel(2+), copper(2+), and zinc(2+). *Biochemistry*, 25:2276, 1986.

[222] J. P. Allen, G. Feher, T. O. Yeates, H. Komiya, and D. C. Rees. Structure of the reaction center from rhodobacter sphaeroides R-26: 5. protein-cofactor (quinones and iron(2+)) interactions. *Proc. Natl. Acad. Sci.*, 85:8487, 1988.

[223] M. H. B. Stowell, T. M. McPhillips, D. C. Rees, S. M. Solitis, E. Abresch,

and G. Feher. Light-induced structural changes in photosynthetic reaction center: implications for mechanism of electron-proton transfer. *Science*, 276:812, 1997.

[224] Q. Xu and M. R. Gunner. Trapping conformational intermediate states in the reaction center protein from photosynthetic bacteria. *Biochemistry*, 40:3232–3241, 2001.

[225] R. H. G. Baxter, N. Ponomarenko, V. Srajer, R. Pahl, K. Moffat, and J. R. Norris. Time-resolved crystallographic studies of light-induced structural changes in the photosynthetic reaction center. *Proc. Natl. Acad. Sci. USA*, 101:5982, 2004.

[226] A. K. Grafton and R. A. Wheeler. Amino acid protonation states determine binding sites of the secondary ubiquinone and its anion in the rhodobacter sphaeroides photosynthetic reaction center. *J. Phys. Chem. B*, 103:5380, 1999.

[227] B. Rabenstein, G. M. Ullmann, and E. Knapp. Electron transfer between the quinones in the photosynthetic reaction center and its coupling to conformational changes. *Biochemistry*, 39(34):10487, 2000.

[228] U. Zachariae and C. R. D. Lancaster. Proton uptake associated with the reduction of the primary quinone Q_A influences the binding site of the secondary quinone Q_B in *Rhodopseudomonas viridis* photosynthetic reaction centers . *Biochim. et Biophys. Acta*, 1505:280–290, 2001.

[229] Susan E. Walden and Ralph A. Wheeler. Protein conformational gate controlling binding site preference and migration for ubiquinone-b in the photosynthetic reaction center of rhodobacter sphaeroides. *J. Phys. Chem. B*, 106:3001–3006, 2002.

[230] P. Ädelroth, M. L. Paddock, L. B. Sagle, G. Feher, and M. Y. Okamura. Identification of the proton pathway in bacterial reaction centers: both protons associated with reduction of Q_B to Q_BH_2 share a common entry point. *Proc. Natl. Acad. Sci.*, 97:13086, 2000.

[231] K. E. McAuley, P. K. Fyfe, J. P. Ridge, N. W. Isaacs, R. J. Cogdell, and M. R. Jones. Structural details of an interaction between cardiolipin and an integral membrane protein. *Proc. Natl. Acad. Sci.*, 96:14706, 1999.

[232] K. E. McAuley, P. K. Fyfe, J. P. Ridge, R. J. Cogdell, N. W. Isaacs, and M. R. Jones. Ubiquinone binding, ubiquinone exclusion, and detailed cofactor conformation in a mutant bacterial reaction center. *Biochemistry*, 39:15032, 1999.

[233] A. J. Chirino, E. J. Lous, M. Huber, J. P. Allen, C. C. Schenck, M. L. Paddock, G. Feher, and D. C. Rees. Crystallographic analyses of site-directed mutants of the photosynthetic reaction center from rhodobacter sphaeroides. *Biochemistry*, 33:4584, 1994.

[234] Q. Xu, H. L. Axelrod, E. C. Abresch, M. L. Paddock, M. Y. Okamura, and G. Feher. X-ray structure determination of three mutants of the bacterial photosynthetic reaction centers from rb. sphaeroides altered proton transfer pathways. *Structure*, 12:703, 2004.

[235] J. Tandori, P. Sebban, H. Michel, and L. Baciou. In rhodobacter sphaeroides reaction centers, mutation of proline L209 to aromatic residues in the vicinity of a water channel alters the dynamic coupling between electron and proton transfer processes. *Biochemistry*, 38:13179, 1999.

[236] A. Kuglstatter, U. Ermler, H. Michel, L. Baciou, and G. Fritzsch. X-ray structure analyses of photosynthetic reaction center variants from rhodobacter sphaeroides: Structural changes induced by point mutations at position L209 modulate electron and proton transfer. *Biochemistry*, 40:4253, 2001.

[237] P. H. McPherson, M. Y. Okamura, and G. Feher. Light-induced proton uptake by photosynthetic reaction centers from rhodobacter sphaeroides r-26. i. protonation of the one-electron

states $D^+Q_A^-$, DQ_A^-, $D^+Q_AQ_B^-$, and $DQ_AQ_B^-$. *Biochim. Biophys. Acta*, 934:348, 1988.

[238] P. H. McPherson, M. Schoenfeld, M. L. Paddock, M. Y. Okamura, and G. Feher. Protonation and free energy changes associated with formation of QBH_2 in native and Glu-L212 - ¿ Gln mutant reaction centers from rhodobacter sphaeroides. *Biochemistry*, 33:1181, 1994.

[239] M. L. Paddock, P. Aedelroth, G. Feher, M. Y. Okamura, and J. T. Beatty. Determination of proton transfer rates by chemical rescue: Application to bacterial reaction centers. *Biochemistry*, 41:14716, 2002.

[240] L. Gerencser and P. Maroti. Retardation of proton transfer caused by binding of the transition metal ion to the bacterial reaction center is due to pK_a shifts of key protonatable residues. *Biochemistry*, 40:1850, 2001.

[241] M. L. Paddock, S. H. Rongey, G. Feher, and M. Y. Okamura. Pathway of proton transfer in bacterial reaction centers: replacement of glutamic acid 212 in the L subunit by glutamine inhibits quinone (secondary acceptor) turnover. *Proc. Natl. Acad. Sci.*, 86:6602, 1989.

[242] P. Sebban, P. Maróti, M. Schiffer, and D. K. Hanson. Electrostatic dominoes: Long distance propagation of mutational effects in photosynthetic reaction centers of rhodobacter capsulatus. *Biochemistry*, 34:8390, 1995.

[243] Z. Zhu and M. R. Gunner. Energetics of quinone-dependent electron and proton transfers in rhodobacter sphaeroides photosynthetic reaction centers. *Biochemistry*, 44:82, 2005.

[244] H. Ishikita, G. Morra, and E. Knapp. Redox potential of quinones in photosynthetic reaction centers from rhodobacter sphaeroides: Dependence on protonation of Glu-L212 and Asp-L213. *Biochemistry*, 42:3882, 2003.

[245] B. Rabenstein, G. M. Ullmann, and E. W. Knapp. Electron transfer between the quinones in the photosynthetic reaction center and its coupling to conformational changes. *Biochemistry*, 39:10487, 1998.

[246] E:. Takahashi and C. A. Wraight. A crucial role for Aspl213 in the proton transfer pathway to the secondary quinone of reaction centers from rhodobacter sphaeroides. *Biochim. Biophys. Acta*, 1020:107–111, 1990.

[247] M. L. Paddock, S. H. Rongey, P. H. McPherson, A. Juth, G. Feher, and M. Y. Okamura. Pathway of proton transfer in bacterial reaction centers: Role of aspartate-L213 in proton transfers associated with reduction of quinone to dihydroquinone. *Biochemistry*, 33:734, 1994.

[248] M. L. Paddock, M. E. Senft, M. S. Graige, S. H. Rongey, T. Turanchik, G. Feher, and M. Y. Okamura. Characterization of second site mutations show that fast proton transfer to Q_B^- is restored in bacterial reaction centers of rhodobacter sphaeroides containing the Asp-L213 -> Asn lesion. *Photosynth. Res.*, 55:281, 1998.

[249] H. Ishikita and E. Knapp. Variation of Ser-L223 hydrogen bonding with the Q_B redox state in reaction centers from rhodobacter sphaeroides. *J. Am. CHem. Soc.*, 126:8059, 2004.

[250] L. M. P. Beekman, R. W. Visschers, R. Monshouwer, M. Heer-Dawson, T. A. Mattioli, P. McGlynn, C. N. Hunter, B. Robert, I. H. M. von Stokkum, R. van Grondelle, and M. R. Jones. Time-resolved and steady-state spectroscopic analysis of membrane-bound reaction centers from rhodobacter sphaeroides: Comparisons with detergent-solubilized complexes. *Biochemistry*, 34:14712, 1995.

[251] E. Nabedryk, J. Breton, R. Hienerwadel, C. Fogel, W. Mantele, M. L. Paddock, and M. Y. Okamura. Fourier

transform infrared difference spectroscopy of secondary quinone acceptor photoreduction in proton transfer mutants of Rhodobacter sphaeroides. *Biochemistry*, 34:14722, 1995.

[252] A. Remy and K. Gerwert. Coupling of light-induced electron transfer to proton uptake in photosynthesis. *Nature Struct. Biol.*, 10:637, 2003.

[253] M. L. Paddock, P. Aedelroth, C. Chang, E. C. Abresch, G. Feher, and M. Y. Okamura. Identification of the proton pathway in bacterial reaction centers: Cooperation between Asp-M17 and Asp-L210 facilitates proton transfer to the secondary quinone (Q_B). *Biochemistry*, 40:6893, 2001.

[254] P. Maróti and C. A. Wraight. Flash-induced H^+ binding by bacterial photosynthetic reaction centers: comparison of spectrophotometric and conductimetric methods. *Biochim. Biophys. Acta*, 934:314, 1988.

[255] P. Maróti and C. A. Wraight. Flash-induced H^+ binding by bacterial photosynthetic reaction centers: comparison of spectrophotometric and conductimetric methods. *Biochim. Biophys. Acta*, 934:314, 1988.

[256] E. Nabedryk, Breton. J., M. Y. Okamura, and M. L. Paddock. Simultaneous replacement of Asp-L210 and Asp-M17 with Asn increases proton uptake by Glu-L212 upon first electron transfer to Q_B in reaction centers from rhodobacter sphaeroides. *Biochemistry*, 40:13826, 2001.

[257] E. Nabedryk, J. Breton, H. M. Joshi, and D. K. Hanson. Fourier transform infrared evidence of proton uptake by glutamate L212 upon reduction of the secondary quinone Q_B in the photosynthetic reaction center from rhodobacter capsulatus. *Biochemistry*, 39:14654, 2000.

[258] A. Mezzetti, E. Nabedryk, J. Breton, M. Y. Okamura, M. L. Paddock, G. Giacometti, and W. Leibl. Rapid-scan fourier transform infrared spectroscopy shows coupling of Glu-L212 protonation and electron transfer to Q_B in rhodobacter sphaeroides reaction centers. *BioChem. Biophys. Acta*, 1553:320, 2002.

[259] C. R. D. Lancaster, H. Michel, B. Honig, and M. R. Gunner. Calculated coupling of electron and proton transfer in the photosynthetic reaction center of rhodopseudomonas viridis. *Biophys. J.*, 70:2469, 1996.

[260] M. S. Graige, G. Feher, and M. Y. Okamura. Conformational gating of the electron transfer reaction qa-q. - ¿ qaqb in bacterial reaction centers of rhodobacter sphaeroides determined by a driving force assay. *Proc. Natl. Acad. Sci.*, 95:11679, 1998.

[261] M. Y. Okamura. Electron transfer in reaction centers from photosynthetic bacteria. In *The 230th ACS National Meeting*, 2005.

[262] M. S. Graige, M. L. Paddock, G. Feher, and M. Y. Okamura. Observation of the protonated semiquinone intermediate in isolated reaction centers from rhodobacter sphaeroides: Implications for the mechanism of electron and proton transfer in proteins. *Biochemistry*, 38:11465–11473, 1999.

[263] M. L. Paddock, P. H. McPherson, G. Feher, and M. Y. Okamura. Pathway of proton transfer in bacterial reaction centers: replacement of serine-L223 by alanine inhibits electron and proton transfers associated with reduction of quinone to dihydroquinone. *Proc. Natl. Acad. Sci.*, 87:6803, 1990.

[264] H. J. Woo, A. R. Dinner, and B. Roux. Grand canonical monte carlo simulations of water in protein environments. *J. Chem. Phys.*, 121:6392–6400, 2004.

[265] M.. Elstner. *Weiterentwicklung quantenmechanischer Rechenverfahren für organische Moleküle und Polymere*. PhD thesis, Universität Paderborn, 1998.

[266] X. Prat Resina and Q. Cui et al. in preparation.

[267] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, V. G. Zakrzewski, J. A. Montgomery, Jr., R. E. Stratmann, J. C. Burant, S. Dapprich, J. M. Millam, A. D. Daniels, K. N. Kudin, M. C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G. A. Petersson, P. Y. Ayala, Q. Cui, K. Morokuma, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. Cioslowski, J. V. Ortiz, A. G. Baboul, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, C. Gonzalez, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, J. L. Andres, C. Gonzalez, M. Head-Gordon, E. S. Replogle, and J. A. Pople. *Gaussian 98*. Gaussian, Inc., Pittsburgh PA, 1998.

[268] S. Mostaghim, M. Hoffmann, P. H. König, Th. Frauenheim, and J. Teich. Molecular force field parametrization using multi-objective evolutionary algorithms. In *Proceedings of the 2004 IEEE Congress on Evolutionary Computation*, pages 212–219, 2004.

[269] W. Humphrey, A. Dalke, and K. Schulten. VMD: visual molecular dynamics. *J. Mol. Graph.*, 14:33, 1996.

[270] E. A. Merrit and D. J. Bacon. Raster3D: Photorealistic molecular graphics. *Meth. Enzymol.*, 277:505, 1997.

[271] O. S. Smart, J. M. Goodfellow, and B. A. Wallace. The pore dimensions of gramicidin A. *Biophys. J.*, 65:2455, 2003.

Own Publications

1. *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, in print. DOI: 10.1021/jp052328q.

2. *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, 9082-9095.

3. *Molecular Force Field Parametrization using Multi-Objective Evolutionary Algorithms*,
   S. Mostaghim*, M. Hoffmann*, P. H. König*, Th. Frauenheim and J. Teich,
   Proceedings of the 2004 IEEE Congress on Evolutionary Computation (2004), 212-219. DOI: 10.1109/CEC.2004.1330859.

4. *Identification of a vibrational Stark shift within an adsorbate layer: NH_3 on Ru(001)*,
   W. Widdra, T. Moritz, K.L. Kostov, P. H. König, M. Staufer, U. Birkenheuer,
   Surface Science, 1999, 430, 558.

* These authors contributed in a similar matter to publication.

Pieces of this thesis were published in (1), (2) and (3).

Acknowledgements

I am most grateful to Prof. Dr. Thomas Frauenheim for accepting me into his group and giving me the opportunity to engage in this field of research. The phrase "for the scientific freedom I was granted" is overused, but it is truly what I have experienced working with Thomas. I have to acknowledge that I was given incredible flexibility regarding which project to pursue next.

This freedom was at least partially made possible through a scholarship within the DFG Graduiertenkolleg "Scientific Computing" at the Universität Paderborn, Paderborn Institute for Scientific Computation (PaSCo), for which I am grateful. The scholarship provided me with the luxury of independence allowing me to sometimes diverge substantially from the overall project of proton transfer in the BRC. The network established by the GK provided my with contacts leading to fruitful discussions and collaborative relationships.

Prof. Dr. Qiang Cui (University of Wisconsin in Madison) played the role of a second advisor. Aside from helping me get started with molecular modelling and CHARMM in particular, Qiang provided me with a steady stream of ideas and comments. The collaboration with him lead to two joint publications so far [116, 124] with more in the makes. I also would like to thank him for all the time I was able to spend in his lab.

Prof. Dr. Martin Karplus (Harvard University, Cambridge, USA) initiated the research project on proton transfer in the bacterial reaction center and made the CHARMM program available. I am grateful for being able to spent some time in his lab for presenting on the progress and discussing issues with him and his group.

These contacts would not have come into being without the help of Jun.-Prof. Dr. Marcus Elstner, to whom I am grateful for arranging the initial contact. Further, he arranged a four week visit to DKFZ for me, letting me meet a number of people engaging in molecular modelling and photosynthesis research.

Michael Hoffmann was a great colleague the last couple of years. Beyond sharing an office we shared a great deal of time working on projects together. We had a great time both in and out of the lab together. Some of our "Let's just see if we can...´´ / "Why don't we try to ...´´ ideas acquired a life of their own. They grew bigger and bigger and we got some nice papers out of them [116, 124, 268].

The CAII computations were carried out together with Nilanjan Ghosh (University of Wisconsin), who provided this real world example for the tests of the proton transfer coordinate. The collaboration lead to a number of suggestions and ideas which brought this development forward in a great way [124].

I met Dr. Sanaz Mostaghim through the GK and was fascinated by the methods she uses and her expertise in multiobjective optimization. Together with Michael we decided to give it a shot and try these methods for force field paramtrization. This collaboration led to a publication [268].

I am grateful to Prof. Dr. G. Matthias Ullmann (Universität Bayreuth) for sharing their list of protonation patterns for the BRC.

Dr. Christof Köhler shared his version of SDFTB with me. Christofs version turned out to be a more stable than the previous version I was working with. Most impor-

tantly, Christof served as a talisman when it came to mysterious computer problems.

In addition, I'm grateful for having stimulating scientific discussion with Dr. Marc Amkreutz, Dr. Christof Köhler (Universität Paderborn), Demian Riccardi, Patti Schaefer, Mark Formaneck, Dr. Guohui Li, Dr. Xavier Prat Resina, Dr. Haibo Yu (University of Wisconsin), Dr. Wei Yang, Dr. Paul Maragakis (Harvard University), Nam Kwango and Ramkumar Rajamani (University of Minnesota).

Most computations were carried out on the Frauenheim group computer cluster as well as the joint cluster with the PC^2, PLING. I would like to thank Michael Sternberg and Zoltán Hajnal for laying the initial groundwork for the computer system. I worked with Christof Köhler and Marc Amkreutz as a team, maintaining and extending the theoretical physics computer system. I apologize for my impatience when it seemed at times that we were full-time system administrators. In the last years Stefan Wippermann, Lukas Kasprowicz and Michael Albrecht, as working student aides, were a great help in taking administrative load off our shoulders.

Further computations were carried out on the clio and suns15k clusters (RRZ Köln) and the Arminius cluster (PC^2).

Often marginalized for their seemingly minor role in success of research are the secretaries of the division although they take a load off the shoulders. My sincere thanks go to Simone Lange, Astrid Canisius and Claudia Jakob for doing a splendid job in this respect.

In addition, I have to thank a couple of people responsible for a great atmosphere both in the lab out of the lab. We had some great outings, events and meetings together: movie nights, barbacue parties, Italian cookouts, a canoing trip. Thanks to all of you for making the time in Paderborn as nice as it was: Michael Hoffmann, Marc Amkreutz, Petra Stammeier, Christof Köhler, Zoltán Hajnal, Marius Wanko, David Heringer, Betty Szûcs and Pierro Altoé.

I am also grateful to the people in Qiang's lab for giving me a warm welcome during the numerous stays, especially Qiang Cui, Demian Riccardi, Mark Formaneck, Nilanjan Ghosh, Xavier Prat Resina and Guohui Li.

I further thank all those unnamed, who in one way or another contributed to the inspiring atmosphere which I enjoyed both at the lab in Paderborn and in Madison.

I am indepted to my parents for cultivating my curiosity and encouraging me to ask questions, even though I drove them and others crazy until I learnt to ask the "right questions". I thank them for their love, understanding and closeness despite gradually growing distances.

Part of my affinity to Madison is the "lid to my pot" which I found during my second visit. I am indepted to my loving wife Jean for supporting me during the last years. Finally I have to thank our daughter Julia for a solid five to six hours of uninterrupted sleep during the night – most of the time – and for giving me that precious smile first thing in the morning.

Colophon

Molecular graphics were generated using a variety of programs: VMD [269], Molscript, Raster3D [270] and hole2 [271].