

Simulation of Long-Range Proton Transfer

—

Development and Application

Zur Erlangung des akademischen Grades eines

DOKTORS DER NATURWISSENSCHAFTEN
(Dr. rer. nat.)

von der KIT-Fakultät für Chemie und Biowissenschaften
des Karlsruher Instituts für Technologie (KIT)
genehmigte

DISSERTATION

von
Thilo Gunter Mast, M. Sc.

1. Referent: Prof. Dr. Marcus Elstner
2. Referent: PD Dr. Andreas-Neil Unterreiner
Tag der mündlichen Prüfung: 17. April 2018



This document is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0): <https://creativecommons.org/licenses/by-sa/4.0/deed.en>

Zusammenfassung

Die archaeele Protonenpumpe Bacteriorhodopsin nutzt Sonnenlicht um einen Protonengradienten über der Zellmembran aufzubauen. Der Mechanismus der Energieumwandlung basiert auf einem gut erforschten Photozyklus. Obwohl schon über 45 Jahre an diesem bioenergetischen Modellsystem geforscht wurde, sind noch immer wenig Informationen über den letzten Schritt dieses Photozyklus vorhanden: den Übergang vom O zum Grundzustand (bR). Es wird jedoch angenommen, dass ein langreichweitiger Protonentransfer den bR Zustand wiederherstellt.

Diese Arbeit zielt darauf ab, weiteren Aufschluss über die mechanistischen Details und thermodynamischen/kinetischen Eigenschaften des O→bR Übergangs zu geben. Um dies zu bewerkstelligen, wurden computergestützte Rechenmethoden als brauchbare Ergänzung zum Experiment verwendet, da die Handhabung von Membranproteinen und die gesetzten Ziele für experimentelle Methoden eine schwierige Aufgabe darstellen.

Zunächst wurden Strukturmodelle mit Hilfe von *enhanced sampling* Methoden für den O, bR und den O* Intermediatzustand erstellt. Für die Simulation der (langreichweitigen) Protonentransferreaktionen wurde ein hybrides QM/MM Set-up basierend auf der semiempirischen Quantenchemie-Methode DFTB3 und dem CHARMM36 Kraftfeld verwendet. Die mit dem Protonentransfer assoziierte Änderung der Gibbs'schen Freien Energie wurde mit Hilfe von Freie-Energie-Techniken in Kombination mit einer fortgeschrittenen Reaktionskoordinate, die auf der *center of excess charge*-Darstellung basiert, aufgeklärt.

Die Ergebnisse weisen darauf hin, dass es sich bei der O→bR Konversion um einen leicht exergonen Prozess handelt, in dem der O* Zustand ein metastabiler Zwischenzustand bildet. Bezüglich des Protonentransfer-Pfads wurde eine interessante Feststellung gemacht: Der langreichweitige Protonentransfer wird über einen Protonen-Loch-/inversen Grotthuss-Mechanismus vollzogen. Zudem wird dieser Ladungstransfer begleitet von der Umorientierung einer funktionell wichtigen, positiv geladenen Arginin Seitenkette sowie der Ausbildung der *proton release group*.

Im Zuge dieser Studien, wurde des Weiteren die Genauigkeit von DFTB3 für Protonentransferreaktionen evaluiert. Die Methode bewies sich als leistungsstark für die Beschreibung dieser Reaktionen mit einer geringfügigen Tendenz für die Unterschätzung von Reaktionsbarrieren. Um diesen Fehler zu korrigieren, der aus der Unterschätzung der kurzreichweitigen Pauli Repulsion in DFTB3 resultiert, wurde die empirische Delta-Pauli Korrektur entwickelt. Allerdings gab es bisher keine Parameter für dieses Model. Daher wurde in dieser Arbeit der erste Parametersatz für CHNO-basierte Molekülsysteme entworfen und an relevanten organischen/biochemischen Systemen evaluiert.

Die Ergebnisse zeigen, dass Delta-Pauli erfolgreich kurzreichweitige Pauli Repulsion in DFTB3 einführt. Dies hat zur Folge, dass Delta-Pauli die Beschreibung von Molekülsystemen bezüglich nichtkovalenter Wechselwirkungen, inter- und intramolekularen Reaktionsbarrieren und Gleichgewichtsgeometrien verbessert.

Abstract

The archaeal proton pump bacteriorhodopsin uses sunlight to build up a proton gradient across the cell membrane. The mechanism of energy conversion is based on a well investigated photocycle. Despite over 45 years of research on this bioenergetical model system, there is still little information available about the last step of the photocycle: the O→ground (bR) state transition. It is merely assumed that a long-range proton transfer recovers the bR state of bacteriorhodopsin.

This work aims to shed further light on the mechanistic details and thermodynamic/kinetic features of the O→bR transition. In order to achieve this, computational methods were used as a viable complement to the experiment since the handling of membrane proteins and the targeted objectives pose a quite involved task to experimental techniques.

Initially, structural models for the O, bR and the O* intermediate state were obtained by employing enhanced sampling molecular dynamics simulations. For the simulation of the (long-range) proton transfer reactions, a hybrid QM/MM setup based on the semiempirical quantum chemistry method DFTB3 and the CHARMM36 force field was used. The change of Gibbs free energy associated with the proton transfer was resolved by employing free energy techniques with an advanced reaction coordinate based on the center of excess charge representation.

The results indicate that the O→bR conversion represents a slightly exergonic process in which the O* state constitutes a metastable intermediate. Concerning the proton transfer pathway, an interesting finding was made: The long-range proton transfer is accomplished via a proton hole/inverse Grotthuss mechanism. Moreover, this charge transfer is accompanied by the reorientation of a functionally important, positively charged arginine side chain as well as by the formation of the proton release group.

In the course of these studies, DFTB3 was furthermore benchmarked for proton transfer reactions. The method proved to be efficient for the description of these reactions with a small tendency to underestimate reaction barriers. In order to correct for this error, which results from missing short-range Pauli repulsion in DFTB3, the empirical Delta-Pauli correction was developed. However, until now, there were no parameters available for this model. Hence, in this work the first parameter set for CHNO-based molecular systems was derived and benchmarked on relevant organic/biomolecular systems.

The results show that Delta-Pauli successfully introduces short-range Pauli repulsion to DFTB3. As a consequence, Delta-Pauli improves the description of molecular systems regarding non-covalent interactions, inter- as well as intramolecular reaction barriers and equilibrium geometries.

Contents

Zusammenfassung	i
Abstract	ii

I Introduction

1 Proton Transfer in the bR O-State	3
1.1 Bacteriorhodopsin	4
1.2 The Photocycle of Bacteriorhodopsin	6
1.3 Open Issues	9
2 A Pauli-Repulsion Correction for DFTB3	11

II Theoretical Background

3 Quantum Chemistry	15
3.1 Density-Functional Theory	16
3.2 Density-Functional Tight-Binding	18
3.3 Treatment of Non-Covalent Interactions in DFTB	22
4 Molecular Mechanics	29
4.1 The CHARMM Force Field	30
4.2 Molecular Dynamics	31
4.3 Quantum Mechanics/Molecular Mechanics	33
5 Enhanced Sampling and Proton Transfer Reactions	35
5.1 Simulated Annealing	37
5.2 (Hamiltonian) Replica-Exchange	38
5.3 Umbrella Sampling	39
5.4 Metadynamics	41
5.5 Proton Transfer Reaction Coordinates	42

III Results

6	Structural Models for the bR/O/O* State	47
6.1	Simulation Setup	47
6.2	Key Differences between the bR/O/O* State	50
6.3	PMF Calculations for the R82 Side Chain	57
6.4	Conclusions	60
7	O→O*: A Direct Proton Transfer	63
7.1	Stability of the Reactant and Product State	63
7.2	The Minimum Free Energy Path	66
7.3	Proton Transfer Pathways over Hydronium and Hydroxide Species	67
7.4	A DFTB3 Benchmark Study: The AspH ⁺ -H ₂ O-Asp System	71
7.5	Conclusions	75
8	O→bR: A Long-Range Proton Transfer	77
8.1	Simulation Setup	77
8.2	Umbrella Sampling for the O*→bR Transition	79
8.3	O→bR: The Complete Reaction Mechanism	84
8.4	Conclusions	85
9	Parametrization and Benchmarking of Delta-Pauli DFTB	87
9.1	A Fitting Approach for Delta-Pauli DFTB	87
9.2	Performance of Delta-Pauli on Training Set Systems	94
9.3	Analysis of the Delta-Pauli Repulsion	95
9.4	Performance of Delta-Pauli on Benchmark Systems	98
9.5	Excursus: Long-Range Corrected DFTB2 for Non-Covalent Interactions	113
9.6	Conclusions	116
10	Summary and Outlook	117

IV Appendix

A	PT Reactions in bR	121
B	Delta-Pauli DFTB	131
C	LC-DFTB2	151
	Bibliography	163
	Abbreviations	179

Part I

Introduction

Proton Transfer in the bR O-State

Sunlight is the primary source of energy and is essential for all life on our planet. Various organisms have trained themselves to derive energy from this source¹. The most prominent example therefore are photoautotrophic organisms like plants, bacteria and algae which directly capture energy from sunlight via chlorophyll-based light-harvesting complexes in order to use it for photosynthesis. Consequently, the photosynthetic process provides energy for the cell and thus drives energy-consuming metabolic processes such as protein synthesis etc. From a thermodynamic point of view this scenario can be seen as a coupling of an exergonic (photosynthesis) to an endergonic process (protein synthesis). This so-called principle of “energy coupling” is ubiquitous in nature and has been perfected in the metabolism of higher life forms.

Beside photosynthesis, cell respiration is the other main energy-yielding process of most organisms. Both processes have in common that they are indirectly coupled to the synthesis of the universal energy carrier of all cells: adenosine triphosphate (ATP). In order to achieve this, the energy which is liberated over a series of electron transfer reactions during photosynthesis or cell respiration is used to drive an uni-directional transport of positively charged hydrogens (protons) across the thylakoid membrane or inner mitochondrial membrane. Via this process, the energy is converted into an electrochemical gradient, also known as protonmotive force, which is applied across the cell membrane and consists of two components: the difference in hydrogen ion concentration and the electric potential difference due to charge separation. The ATP synthase, another vectorial ion transporter dissipates this protonmotive force and drives with it the endothermic reaction of adenosine diphosphate with inorganic phosphate to produce ATP. As a result, the energy is stored in the pyrophosphate bond. This whole concept is known as the “chemiosmotic theory”, which was first proposed by Peter Mitchell in 1961² and who was rewarded with the Nobel prize in 1978. The usage of a molecular energy carrier like ATP ensures that the energy provision is not bound to a single compartment like chloroplasts (photosynthesis) or mitochondria (respiration) but can be delivered to at least, other parts within the cell in order to drive endergonic processes. Furthermore, the energy is temporarily stored and available also when the primary energy source (e.g. sun) is absent. Due to these properties, ATP is the main energy currency in the cell.

A special form of photoautotrophy, which is not based on chlorophyll as the light-harvesting complex, can be found in the cell membranes of microorganisms from the class of Halobacteria.

1.1 Bacteriorhodopsin

Unlike the name suggests, Halobacteria belong to the Archaea rather than to the Bacteria domain³. These extremophilic organisms prosper under high salt concentrations like in the Great Salt Lake or the Dead Sea. As a result, parts of those lakes are colored purple-red due to the presence of Halobacteria.

An explanation for this was given in 1971, when Oesterhelt and Stoeckenius⁴ discovered in one of such halophilic archaea, *Halobacterium halobium*, the integral membrane protein bacteriorhodopsin (bR). They showed that this protein consists of an opsin-like part bound to a photoactive retinal molecule, which exhibits an absorption maximum at 570 nm. This chromophore is responsible for the discovered purple patches of the cell membrane⁵ and hence for the coloring of the lakes.

A few years later, in 1973, the function of bR was elucidated: The archaeal protein serves as a light-harvesting complex, which uses the energy of light to pump protons across the cell membrane⁶. During this vectorial transport a protonmotive force results, which drives ATP synthesis and in turn the energy production of the cell⁷⁻¹⁰. According to this, the photosynthetic process taking place in bR bears analogy to the photosynthesis in the chloroplasts.

The deep interest and the experimental advantages of bR led to its first low-resolution crystal structure shortly afterwards, in 1975¹¹. This fact makes bR the first membrane protein whose structure was resolved by any experimental technique. According to the crystal structure, bR occurs as a trimeric unit with a 3:1 peptide:lipid-ratio in the membrane (see Figure 1.1A). The functional monomer of bR, depicted in Figure 1.1B, consists of 248 amino acids with a molecular weight of 24 kDa. And there are seven α -helices, which penetrate the membrane layer almost perpendicularly.

In the years that followed, bR advanced to a model system for structural biology and bioenergetical transport mechanisms. This is mainly due to the proteins theoretical simplicity and experimental advantages^{12,a}. Nowadays, the protein data bank RCSB¹³ contains 131 bR crystal structures, mostly based on x-ray diffraction with high-resolutions of up to 1.43 Å¹⁴. With respect to these data, the photoactive retinal molecule comprises a conjugated polyene chain with a β -ionone ring. Furthermore, this molecule is linked via a protonated Schiff base¹⁵ to the ϵ -amino group of K216^{4,16,17} and thereby separates the proteins cytoplasmic and extracellular part. Neutron diffraction studies¹⁸ and crystallographic B-factors¹⁹⁻²¹ indicated that the cytoplasmic part is more flexible than the rigid extracellular part. This fact is a result of the amino acid and water composition of the protein: The cytoplasmic part is dominated by hydrophobic amino acids, whereas in the extracellular

^ae.g. bR tends to form well-ordered two-dimensional crystals.

part mainly charged and polar amino acids dominate, which form a network of hydrogen bonds together with several water molecules, making this region more rigid¹².

The proton pumping mechanism was studied extensively with various different (also time-resolved) spectroscopic methods in combination with experimental techniques like cryotrapping. These studies showed that the pumping mechanism is initiated by exposure of the retinal molecule to light. This primary event triggers a sequence of large structural rearrangements in the protein, leading to several different intermediate states before the initial state, the so-called ground (bR) state, is recovered. The sequence of the individual intermediate states is known as the photocycle. Despite the fact that most of these intermediate states are very well investigated, until today there is no consensus on the complete reaction mechanism. Reasons for that are manifold and will not be discussed in this work. Instead, the reader is referred to refs. 12,22,23. However, a generally accepted photocycle scenario will be outlined in the next section.

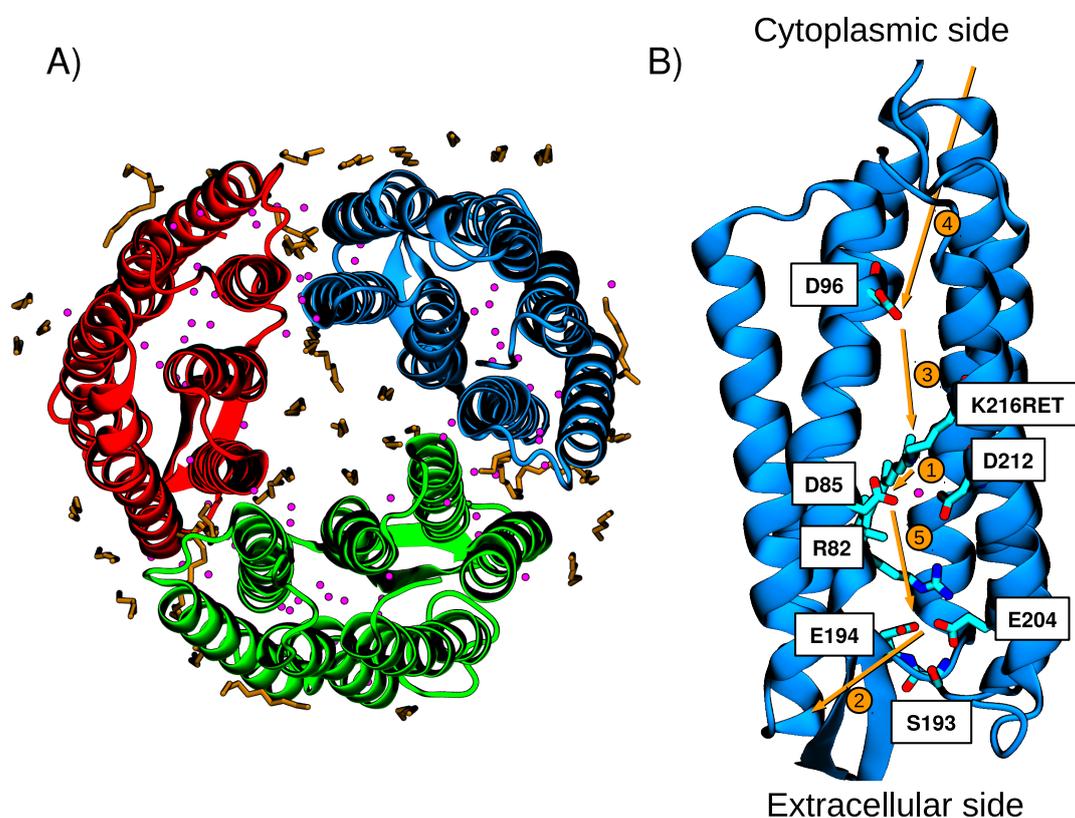


Figure 1.1: Three-dimensional structures of the bR state. A: Top view on bR in its naturally occurring trimeric form. Lipids are depicted in ochre and crystal waters in pink. Based on PDBID 4XXJ²⁴. B: Side view on the bR monomer. Functionally important amino acids are shown and labeled. Key proton transfer steps are labeled from 1 to 5 and are further described in the text. Based on PDBID 5B6V²⁵.

1.2 The Photocycle of Bacteriorhodopsin

bR→K Transition

In the bacteriorhodopsin ground state (bR) the Schiff base of the retinal is facing the extracellular side. This configuration is stabilized by a hydrogen-bonded network (HBN), which is formed by the protonated Schiff base together with three water molecules W400, W401, W402 and the aspartate side chains of D85 and D212^{20,21,26}. The photocycle of bR and therewith the proton pumping mechanism is initiated, when a photon with a wavelength of 570 nm^b is absorbed by the retinal. As a consequence, the chromophore is excited from the electronic ground state S_0 into its first excited state S_1 ²⁷. During relaxation, an isomerization around the $C_{13}=C_{14}$ bond converts the retinal from its all-*trans* to its 13-*cis*,15-*anti* form. This process takes place on a timescale of 4 ps and leads to the K state²⁸⁻³⁰ (see Figure 1.2B).

K→L Transition

The red-shifted K state exhibits a twisted configuration of the isomerized retinal, where the $N\zeta-H$ bond of the Schiff base is reoriented towards the cytoplasmic side. Furthermore, the hydrogen bond to the structurally important water W402 is weakened/broken so that the Schiff base approaches D85³¹. This strained configuration is a result of the stored photon excess energy. In fact, the K state can be seen as the intermediate with the highest free energy which drives the subsequent reactions of the photocycle¹². The mechanistic details of the K→L transition are still under discussion, because of the strong dependence on experimental conditions and probably spectroscopically silent intermediates³²⁻³⁴. After approximately 1 μ s the L state is formed.

L→M₁ Transition

It is assumed that the strain on the retinal chromophore is reduced and the hydrogen bond to W402 reestablished such that the first proton transfer (PT) step is prepared. During the L→M₁ transition, which lies on a timescale of roughly 40 μ s the Schiff base is deprotonated with D85 as the proton acceptor. However, there is no consensus about the mechanism of the PT. Proposed scenarios involve a direct PT as well as an indirect PT over the W402 water³⁵⁻³⁹. It was also proposed that W402 is first deprotonated, leading to a proton hole which then deprotonates the Schiff base⁴⁰. The PT releases further strain on the retinal such that the chromophore and the D85 residue are drawn apart making the PT irreversible¹². Consequently, the change in the protonation pattern directly affects the R82 side chain and induces a swing movement of this amino acid towards the E194/E204 and away from D85/D212⁴¹⁻⁴³. Clemens et al.⁴⁴ observed this movement also in a molecular dynamics (MD) simulation. Furthermore, on the basis of Potential of Mean Force (PMF) calculations they could show that the extracellular orientation of R82 is preferred by roughly 0.9 kcal/mol over the cytoplasmic-oriented one. According to their results, the extracellular-oriented R82

^bThis is equivalent to an energy of 2.175 eV or 50.2 kcal/mol.

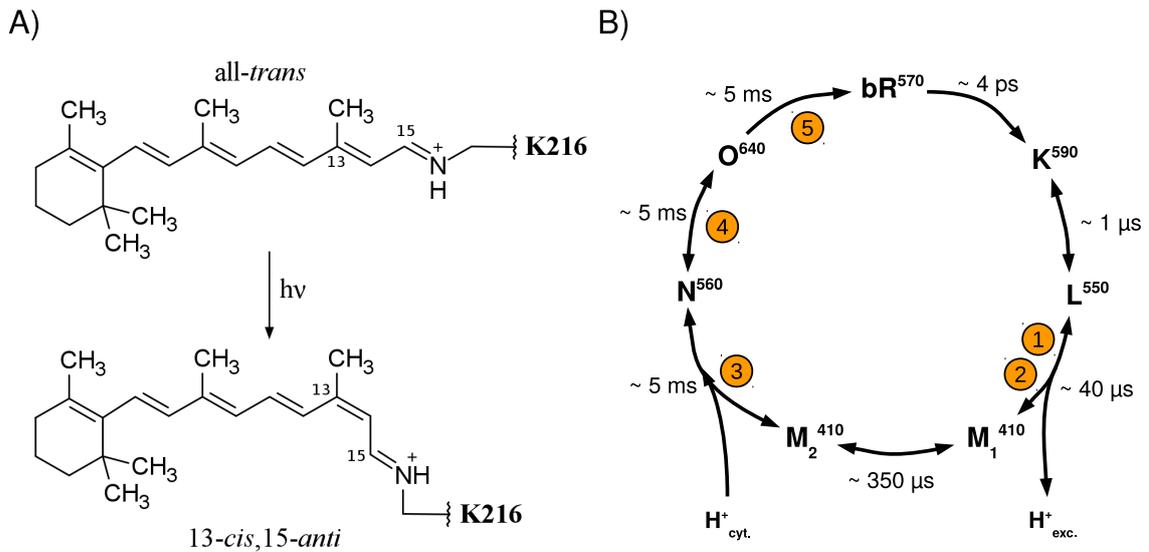


Figure 1.2: The different retinal configurations and the bR photocycle. A: The absorption of a photon triggers the isomerization of the all-*trans* retinal in its 13-*cis*,15-*anti* form. The retinal is linked via a protonated Schiff base to the protein part. B: Schematic representation of the bR photocycle. Important intermediates, their absorption maxima and approximate lifetimes are shown. The key proton transfer steps are labeled from 1 to 5 (cf. Figure 1.1B).

configuration allows the influx of additional water molecules from the bulk. The R82 side chain movement, in turn, changes the electrostatics in the protein environment close to the extracellular side¹⁹. As a result, a proton is released at the extracellular side coming from the so-called proton release group (PRG). The identity of this group is still debated. Popular suggestions are that this group is the protonated amino acid E204⁴⁵, a complex of E204 and E194⁴⁶ or a protonated water cluster between R82 and E204/E194⁴⁷. With the help of hybrid quantum mechanics/molecular mechanics (QM/MM) simulations it has been shown that the last option is energetically very unfavorable, because PMF calculations indicated that there is no stable conformation for a protonated water cluster^{48,49}. According to Goyal et al.⁴⁹ a complex with a shared proton between E194/E204 and stabilized by S193 is approximately 15 kcal/mol more favorable than the protonated water cluster.

M₁→M₂ Transition

In the strong blue-shifted M₁ state, the deprotonated retinal Schiff base is still in a rather extracellular- than cytoplasmic-oriented configuration. However, since bacteriorhodopsin is a vectorial proton pump it must be ensured that the Schiff base can be reprotonated by a proton coming from the cytoplasmic side. Therefore, the Schiff base has to reorient in a so-called switch step which is happening during the M₁→M₂ transition²³. This spectroscopically silent transition happens on a time scale of 350 μs⁵⁰. It is believed that the switch mechanism induces large conformational changes in the cytoplasmic protein region⁵¹.

M₂→N Transition

The N ζ group of the Schiff base is reprotonated by D96 during the M₂→N transition^{52–58}. Due to the large distance between the proton acceptor and donor (approximately 12Å) and a transition time of ca. 5 ms, it is assumed that the PT is accomplished over a transient water chain in a Grotthuss-like mechanism⁵⁹.

N→O Transition

During the approximately 5 ms long N→O transition a hydrophobic plug in the cytoplasmic region gets opened, allowing the influx of a proton from the cytoplasmic water phase which then reprotonates D96^{53,55}. Probably at the same time the retinal thermally re-isomerizes back in its all-*trans* configuration⁶⁰. The synchronism of both processes suggests that they are coupled to each other⁶¹.

O→bR Transition

Compared to the early half of the photocycle, the experimental data for the O state as well for the transition back to the bR state is sparse: Currently, the RCSB database contains four O-like crystal structures, but these are all based on bR mutants. It is assumed that the newly formed all-*trans* retinal adopts a twisted conformation⁶² in the red-shifted O state. Furthermore, structural rearrangements of the protein helices lead to a more open conformation of the extracellular side. For the return into the bR state, which is assumed to happen on a time scale of 5 ms⁶³, D85 has to be deprotonated and the PRG protonated. Moreover, the change in the protonation pattern should trigger the upswing movement of the R82 side chain back in its initial position. Since the low initial pK_a of D85 and the high initial pK_a of the PRG are recovered during this process, it is expected that this reaction is strongly downhill^{12,23}. Furthermore, the deprotonation of D85 is assumed to be the rate-limiting step in this process⁶⁴. However, the D85 amino acid is spatially separated by roughly 12–14 Å from the E204/E194 amino acids, and the positively charged R82 side chain lies directly in between. The appearance of a new C=O stretch band in time-resolved Fourier transform infrared spectra gave the hint that D212 is transiently protonated during the O→bR reaction^{65,66}, leading to an intermediate state O*. Minimum energy path (MEP) calculations, based on the conjugate peak refinement (CPR) method⁶⁷ showed that an indirect PT from D85 to D212 over water molecules is indeed kinetically feasible with a reaction barrier between 4 – 11 kcal/mol^c. Furthermore, these results revealed the endothermic character of this reaction leading to a metastable O* state. This in turn led to the conclusion that D212 only serves as a proton-stabilizing intermediate in a larger PT. The dominant opinion in the bR community hence suggest that like in the M₂→N transition, a long-range PT takes place in which the D85 proton is transferred over a water wire and via D212 down to the PRG. However, CPR calculations⁶⁸ have shown that a Grotthuss-like PT from D85→E204 exhibits a very high reaction barrier of 21.9 kcal/mol

^cDepending on the reaction pathway and the number of water molecules in the pathway.

and is strongly endothermic ($\Delta E = 10.0 \text{ kcal/mol}$)^d. It is worth mentioning that the O-like D85S mutant⁷² exhibits the ability to pump chloride ions from the extracellular towards the cytoplasmic side^e.

To sum up, in the bR photocycle the collected photon energy of ca. 50 kcal/mol is gradually converted into a protonmotive force by releasing one proton to the extracellular side.

1.3 Open Issues

Despite over 45 years of research on bR, the exact mechanism of the photocycle still is not revealed. Especially structural, mechanistic and energetic details of the late O state and its conversion back to the bR state remain elusive.

The problem is the sparse information about the important degrees of freedom, which drive the O→bR transition. Hence, it is necessary to find the most stable conformations of the O and O* state and then compare them to the bR structure in order to identify any possibly important degrees of freedom. One such, beside the long-range PT, has to be the swing movement of the R82 side chain. However, it is not clear at what stage in the O→bR reaction this side chain reorients towards the cytoplasmic side. It is also not obvious, if there is enough cavity water for the assumed long-range PT available between the D85/D212 and E194/E204 site. For the PT the cavity waters have to form a transient water chain from D85 to the PRG, which could be hindered by the R82 side chain which lies in between.

Regarding the O→O* transition, the exact pathway of the PT was not resolved because the mentioned MEP calculations⁶⁷ implied a predefinition of the reaction pathway (PT over one or two etc. water molecules), and neglected the equilibration of the environment. However, a PT entails a change in the protonation pattern, which could have drastic effects in a polar region like this. Consequently, if structural rearrangements arise due to this transfer one would not notice them.

With respect to the energetics of the O→bR transition, the situation looks even worse: The two proposed pathways seem to be either endothermic (Grotthuss pathway over water wire) or kinetically unfeasible due to the high reaction barrier (pathway over R82 side chain)⁶⁸. Moreover, it is well known that the D85/D212 site is electrostatically coupled to the E194/E204 site. Since the long-range PT alters the protonation state of these sites, it is also probable that this event triggers further reactions, which are not resolved by MEP calculations. Furthermore, entropic and solvent effects, which are neglected in MEP calculations, could also play important roles in the O→bR transition.

^dA yet other scenario which is proposed, suggests the R82 side chain as simultaneous proton acceptor/donor in the D85→PRG PT⁶⁹. Nevertheless, R82 mutants are still able to pump protons (even though with reduced efficiency)^{70,71} and MEP calculations yield a too high reaction barrier of roughly 36 kcal/mol for this pathway⁶⁸.

^ehR, which is another archaeal rhodopsin-based protein closely related to bR, pumps chloride ions in its native form⁷³.

Therefore, this work aims to:

- Propose structural models for the bR, O and O* states. Thereby the main stabilizing interactions, their energetics and the most significant differences between the models will be figured out for the active site between D85 and the PRG. Furthermore, the water density in this region will be in the focus. These analyses are necessary in order to detect the important degrees of freedom of the O→bR transition.
- Predict Gibbs free energies for the O→O* as well as for the O→bR transition. Thereby the PT pathways (which should not be assumed *a priori*) and the response of the environment (structural rearrangements etc.) are of special interest. The free energies in turn can then be used in order to
- Propose a complete, kinetically feasible and exergonic reaction mechanism for the O→bR transition.

As already mentioned, experimental researches struggle with the correct description of the O state. As a consequence, it is difficult to resolve the stated issues with these methods. Nevertheless, MD simulations won fame especially for the study of membrane proteins which are experimentally hard accessible. In order to give answers to the stated issues, various MD simulation techniques will be employed in this work. Classical force-fields are known to efficiently predict the time-dependent evolution of large molecular complexes, consisting of multiple thousands of atoms. However, since the project implies the simulation of chemical reactions, in which the electronic structure rearranges, a quantum mechanical (QM) description is inevitable but much too expensive for the whole system. Consequently, the multi-scale QM/MM approach will be applied where the small, reactive region of the molecular system is described by quantum chemistry and the rest by classical Newtonian mechanics^f. In this work, the QM region will be treated with the semiempirical density functional tight-binding (DFTB) method. The semiempirical character of DFTB allows efficient sampling of large QM regions, which is necessary in order to predict free energies. However, the accuracy of this method suffers sometimes from its approximative nature. Especially the description of non-covalent interactions like Pauli repulsion is insufficient and need to be corrected, for example with an auxiliary model. The parametrization of one such model, which should improve the Pauli repulsion in DFTB represents the second part of this thesis.

^fIn 2013, Karplus, Levitt and Warshel received the Nobel prize in Chemistry for the development of the QM/MM approach.

A Pauli-Repulsion Correction for DFTB3

DFTB is a semiempirical quantum chemistry method, which is derived from Kohn-Sham density functional theory (DFT) by a Taylor series expansion⁷⁴. While the application of the original DFTB model^{75,76} was restricted to the solid state, the self-consistent charge (SCC) models, DFTB2⁷⁷ and the current model DFTB3^{78–81} gained reputation in the field of soft matter. Especially in the framework of QM/MM calculations, where large QM regions of up to hundreds of atoms or long sampling times up to multiple nanoseconds are the rule, the efficient DFTB method is extensively used⁸². Consequently, DFTB also found application to various PT reactions occurring in proteins, for example in bR^{35,48,49,67,68}, cytochrome c oxidase⁸³, carbonic anhydrase^{84–86} etc. The speedup effect of DFTB results from approximations like the use of a minimal basis set and exclusive treatment of valence electrons. Furthermore, the computationally expensive Hamiltonian and overlap matrices contain one- and two-center contributions only. Therefore, they can be parametrized on basis of the generalized gradient approximation (GGA)-DFT functional PBE⁸⁷ in advance such that they just need to be read in during the calculation. As a result, DFTB runs two to three orders of magnitude faster than a conventional DFT calculation while still yielding results comparable to a GGA-DFT method employing a double- ζ basis set⁸¹.

Nevertheless, the approximative character of DFTB and also inherited drawbacks from the PBE functional are responsible for this method hardly reaching the chemical accuracy (1.0 kcal/mol). Especially the description of non-covalent interactions like attractive dispersion and Pauli repulsion is a delicate issue. One has to mention that the insufficient description of these interactions is not a DFTB-specific problem as *ab initio* methods suffer from this difficulty also. The reason for this can be traced back to the inadequate description of electron correlation effects in these methods⁸⁸. However, these interactions influence important molecular properties like geometries, reaction energies etc. For example, the stability of the π -stacked benzene dimer (see Figure 2.1A) or base pair stacking in nucleic acids is mainly a result of mid- to long-range attractive dispersion interactions between the π -systems. On the other hand, at short interatomic distances the Pauli repulsion becomes important for the energetics. Hence, reaction barriers like the rotational barrier of

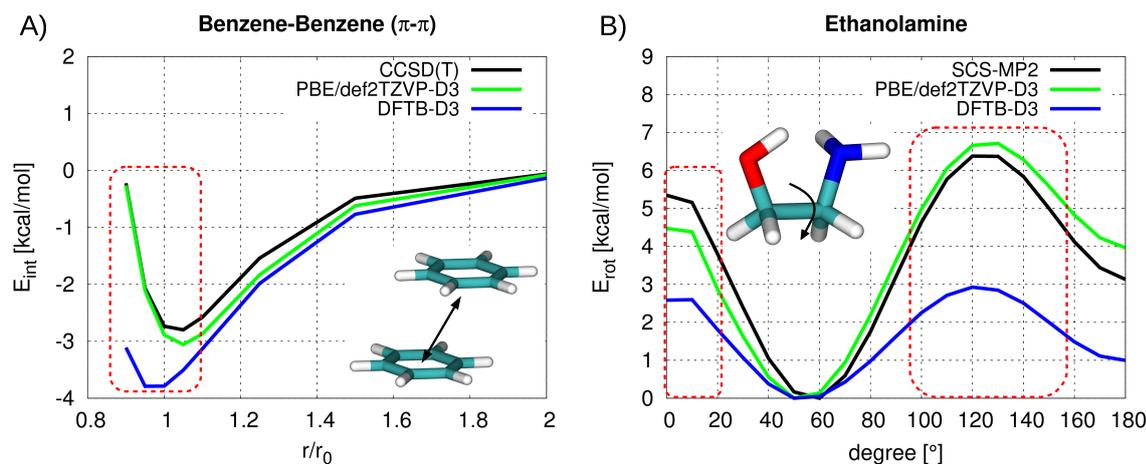


Figure 2.1: Interaction energy of the benzene dimer (A) and rotational barrier of ethanolamine (B). The data was calculated with a reference method (CCSD(T)⁹⁸ or SCS-MP2), PBE-DFT-D3 and DFTB-D3. A: The D3 correction ensures for the correct description of the dispersion-dominated mid- to long-range part ($r/r_0 > 1.1$). However, for the short-range Pauli-repulsive dominated part (red box) DFTB-D3 underestimates this contribution and consequently deviates from the reference and its well performing parent method PBE-DFT-D3. B: DFTBs underestimation of Pauli repulsion leads beside other effects to a continuous underestimation (red boxes) of rotational barriers.

ethanolamine (see Figure 2.1B) are mainly dictated by this interaction.

In order to account for non-covalent interactions, quantum chemistry methods which aim at efficiency, use empirical corrections. As a result, the D3 model by Grimme⁸⁹ advanced in the recent years to the most popular and efficient dispersion correction model. It was parametrized for most quantum chemical methods including DFTB3^{90,91}.

However, a correction for the short-range Pauli repulsion is still missing such that this interaction is generally underestimated in DFTB. This is mainly because of two reasons: Already the PBE functional underestimates this interaction such that DFTB inherits this deficiency due to parametrization. Secondly, this can be seen as a result of the minimal basis set approximation. Consequently, energetic properties like reaction barriers are generally underestimated. This deficiency also affects geometrical properties such that DFTB describes condensed matter generally as too dense⁹², leading for example to overcoordinated⁹³ complexes, a wrong radial distribution function for water^{94,95}, voids in water clusters⁹⁶ etc.

Consequently, there is a need for a Pauli repulsion correction in DFTB. Recently, Maximilian Kubillus⁹⁷ proposed one such empirical correction model for DFTB3, named Delta-Pauli. In this work the first parameter set for Delta-Pauli will be derived and benchmarked on organic and bioorganic systems.

Part II

Theoretical Background

Quantum Chemistry

In order to resolve the electronic structure of a molecular system, the corresponding Schrödinger equation has to be solved. In most cases, atoms or molecules are investigated which exhibit no time-dependent interactions^a, so that the stationary Schrödinger equation is solved:

$$\hat{H}\Psi(\mathbf{r}_1 \dots \mathbf{r}_n, \mathbf{R}_1 \dots \mathbf{R}_N) = E\Psi(\mathbf{r}_1 \dots \mathbf{r}_n, \mathbf{R}_1 \dots \mathbf{R}_N) \quad (3.1)$$

where \hat{H} is the time-independent Hamilton operator, Ψ represents the many-body wave function for a system of n electrons with coordinates \mathbf{r} and N nuclei with coordinates \mathbf{R} , and E is the corresponding eigenvalue to \hat{H} representing the total energy of the system. This many-particle problem may be simplified via introduction of the Born-Oppenheimer approximation⁹⁹. The key message of this approximation is that the movement of the electrons can be decoupled from those of the nuclei^b. As a consequence, the electronic Hamiltonian depends only on the electronic coordinates \mathbf{r} , while the coordinates of the nuclei \mathbf{R} enter solely as parameters:

$$\begin{aligned} \hat{H} &= \hat{T}_e + \hat{V}_{ee} + \hat{V}_{eN} \\ &= -\frac{1}{2} \sum_i^n \nabla_i^2 + \sum_i^n \sum_{j>i}^n \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} - \sum_i^n \sum_A^N \frac{Z_A}{|\mathbf{r}_i - \mathbf{R}_A|} \end{aligned} \quad (3.2)$$

\hat{T}_e represents the kinetic energy of the electrons, \hat{V}_{ee} is the electron-electron interaction, \hat{V}_{eN} describes the interaction of the electrons with the nuclei.

Although the Born-Oppenheimer approximation reduces the complexity of the problem, there is still the many-body interaction of the electrons that makes an analytical and hence exact solution of the Schrödinger equation unique to one-electron systems. Nevertheless, there is a recipe for approaching the ground state wave function Ψ_0 with which all quantum-chemical methods deal with: the variational principle. It states that the energy

^ae.g. no interaction with an electric or magnetic field.

^bThis approximation is justified by the fact that the nuclei are much heavier and hence slower than the electrons. Thus, the electrons move from their point of view in a constant potential of the nuclei. Furthermore, the nuclear kinetic energy term is assumed to be independent of the electrons and the correlation in the electron-nuclear potential is lost.

E_g corresponding to a trial wave function Ψ_g is an upper bound to the true ground state energy E_0 :

$$\langle \Psi_g | \hat{H} | \Psi_g \rangle = E_g \geq E_0 = \langle \Psi_0 | \hat{H} | \Psi_0 \rangle \quad (3.3)$$

From this point start the two main classes of quantum chemistry: Hartree-Fock (HF) theory and DFT. We will continue with the latter method, because it was used in this work and furthermore is the basis for the DFTB method.

3.1 Density-Functional Theory

The basis of DFT provides the replacement of the many-body wave function $\Psi(\mathbf{r}_1 \dots \mathbf{r}_n)$ by the electron density $\rho(\mathbf{r})$:

$$\Psi(\mathbf{r}_1 \dots \mathbf{r}_n) \rightarrow \Psi[\rho(\mathbf{r})] \quad (3.4)$$

As a consequence, the system is described by a function of three instead of a function of $3N^c$ coordinates. The use of the electron density as the central variable is legitimized by the theorems of Hohenberg and Kohn¹⁰⁰, which state that the ground state and its properties like the total energy E_0 are a unique functional of the ground state electron density $\rho_0(\mathbf{r})$:

$$E_0 = E[\rho_0(\mathbf{r})] \quad (3.5)$$

The main goal of all DFT methods is to approach the exact functional which connects the electron density $\rho_0(\mathbf{r})$ to the energy E_0 . As an ansatz, the total energy functional can be decomposed into its contributions:

$$E[\rho] = T[\rho] + V_{\text{ext}}[\rho] + J[\rho] + E_{xc}[\rho] \quad (3.6)$$

with $T[\rho]$ as the kinetic energy of the electrons, $V_{\text{ext}}[\rho]$ as the external potential^d, $J[\rho]$ as the classical electron-electron Coulomb repulsion and $E_{xc}[\rho]$ as the exchange-correlation energy. It is worth mentioning that the exchange-correlation term contains all the non-classical contributions like self-interaction correction, exchange and electron correlation effects.

The first attempts to construct this energy functional were made by Thomas¹⁰¹, Fermi¹⁰² and Dirac¹⁰³. However, due to the pure density functional approach the description of the kinetic energy term was insufficient. In order to gain a correct representation of the kinetic energy, Kohn and Sham¹⁰⁴ reintroduced molecular orbitals (MOs) to DFT. They assumed that the orbitals could be used to obtain the exact kinetic energy of an auxiliary system of non-interacting electrons, representing the same electron density as the real, interacting system. The difference between the kinetic energy of both systems, the classical self-interaction correction as well as all non-classical contributions like exchange and electron correlation effects is then lumped together into the $E_{xc}[\rho]$ term. With this, the Kohn-Sham expression of the total energy takes the form:

$$E[\rho] = \sum_i^n \left\langle \psi_i \left| -\frac{1}{2} \nabla^2 \right| \psi_i \right\rangle + \int V_{\text{ext}}(\mathbf{r}) \rho(\mathbf{r}) d\mathbf{r} + \frac{1}{2} \iint \frac{\rho'(\mathbf{r}) \rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r} d\mathbf{r}' + E_{xc}[\rho(\mathbf{r})] + V_{NN} \quad (3.7)$$

^cDue to the N-electron wave function.

^dEnergy of the electrons in the external field of the nuclei.

An analytical solution can be found for all terms in the upper equation 3.7, except for the E_{xc} term, which needs to be approximated by a functional. The total energy can then be minimized by applying the variational principle under the orthonormality constraint $\langle \psi_i | \psi_j \rangle = \delta_{ij}$. This leads to an eigenvalue problem consisting of a set of one-electron equations, which are known as the Kohn-Sham equations:

$$\hat{h}_i \psi_i = \epsilon_i \psi_i \quad (3.8)$$

with ϵ_i as the energy eigenvalue of electron i and \hat{h}_i as the corresponding single electron Hamiltonian defined as:

$$\hat{h}_i = -\frac{1}{2} \nabla_i^2 - \sum_A \frac{Z_A}{|\mathbf{r}_i - \mathbf{R}_A|} + \int \frac{\rho(\mathbf{r}')}{|\mathbf{r}_i - \mathbf{r}'|} d\mathbf{r}' + V_{xc} \quad (3.9)$$

The exchange-correlation potential V_{xc} is defined as the derivative of the exchange-correlation energy E_{xc} with respect to the electron density:

$$V_{xc} = \frac{\delta E_{xc}[\rho(\mathbf{r})]}{\delta \rho(\mathbf{r})} \quad (3.10)$$

The approximation of the unknown exchange-correlation functional E_{xc} is crucial for the performance of a DFT method and there are four main classes in modern DFT, shown in Table 3.1, which treat this quantity differently. At this point, it is important to mention that if the exact exchange-correlation potential V_{xc} was known, the Kohn-Sham approach could take electron correlation effects into account exactly and hence would lead to the exact energy of the system. Since the Kohn-Sham operator \hat{h}_i itself depends on the electron density $\rho(\mathbf{r})$, the set of one-electron equations has to be solved in a self-consistent manner. The computational cost of this procedure is cheaper than solving the Hartree equations

Table 3.1: Classes of exchange-correlation DFT functionals.

Class	Description	Popular Func.
local density approximation (LDA)	E_{xc} depends on the local $\rho(\mathbf{r})$.	SPW92 ¹⁰⁵
generalized gradient approximation (GGA)	E_{xc} depends on the local $\rho(\mathbf{r})$ but also on its local gradient $\nabla\rho(\mathbf{r})$.	PBE ⁸⁷ BLYP ^{106,107}
meta-GGA	The GGA is extended by the dependence on higher-order derivatives of $\rho(\mathbf{r})$. Alternative approaches are using the orbital kinetic energy density.	TPSS ¹⁰⁸
hybrid	E_{xc} is described as a functional composed of HF and DFT exchange.	B3LYP ^{107,109–111}

in HF theory^e. Similarly as in HF, nearly all DFT applications make use of the linear combination of atomic orbitals (LCAO) approach in order to expand the Kohn-Sham MOs ψ_i in a set of L atomic basis functions η_μ , weighted by the coefficients c_μ^i :

$$\psi_i = \sum_{\mu}^L c_{\mu}^i \eta_{\mu} \quad (3.11)$$

Due to their computational cost the application of DFT methods in QM/MM molecular dynamics simulations is still limited to a small system size of less than 100 atoms and to a time regime of picoseconds. This fact represents a serious constraint for the study of biophysical processes where sufficient sampling over nanoseconds to microseconds is essential. As a consequence, fast semiempirical methods are still demanded and further developed. Therefore, the following section introduces an approximate DFT method, which is able to treat larger QM systems on a longer timescale than pure DFT methods.

3.2 Density-Functional Tight-Binding

DFTB has its roots in solid state physics. Like for semiempirical quantum chemical methods, the computational efficiency is increased by circumventing time-consuming calculations via parameterization and introduction of several approximations.

In order to derive the DFTB formalism from the DFT total energy, the electron density $\rho(\mathbf{r})$ is expressed as the superposition of atomic reference densities $\rho_0(\mathbf{r})$ and their fluctuations $\delta\rho(\mathbf{r})$ ¹¹²:

$$\rho(\mathbf{r}) = \sum_a \rho_a(\mathbf{r}) + \delta\rho(\mathbf{r}) \quad (3.12)$$

Equation 3.12 is then inserted in the DFT total energy (see equation 3.7) and the exchange-correlation energy term $E_{xc}[\rho_0(\mathbf{r}) + \delta\rho(\mathbf{r})]$ is expanded in a Taylor series up to the third order around the reference density $\rho_0(\mathbf{r})$. All terms containing linear density fluctuations $\delta\rho(\mathbf{r})$ vanish due to the variational theorem and we end up with the DFTB total energy:

$$\begin{aligned} E^{\text{DFTB}} &= E_{xc}[\rho_0(\mathbf{r})] + E_{NN} - \int V_{xc}[\rho_0(\mathbf{r})]\rho_0(\mathbf{r})d\mathbf{r} - \frac{1}{2} \iint \frac{\rho_0(\mathbf{r})\rho_0(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}d\mathbf{r}' \\ &+ \sum_i^{\text{occ}} \left\langle \psi_i \left| \underbrace{-\frac{1}{2}\nabla^2 + V_{ext} + \int \frac{\rho_0(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' + V_{xc}[\rho_0(\mathbf{r})]}_{\hat{H}^0[\rho_0(\mathbf{r})]} \right| \psi_i \right\rangle \\ &+ \frac{1}{2} \iint \left(\frac{\delta^2 E_{xc}[\rho_0(\mathbf{r})]}{\delta\rho(\mathbf{r})\delta\rho(\mathbf{r}')} + \frac{1}{|\mathbf{r} - \mathbf{r}'|} \right)_{\rho_0(\mathbf{r})} \delta\rho(\mathbf{r})\delta\rho(\mathbf{r}') d\mathbf{r}d\mathbf{r}' \\ &+ \frac{1}{6} \iiint \frac{\delta^3 E_{xc}[\rho_0(\mathbf{r})]}{\delta\rho(\mathbf{r})\delta\rho(\mathbf{r}')\delta\rho(\mathbf{r}'')} \Big|_{\rho_0(\mathbf{r})} \delta\rho(\mathbf{r})\delta\rho(\mathbf{r}')\delta\rho(\mathbf{r}'') d\mathbf{r}d\mathbf{r}'d\mathbf{r}'' \\ &= E^0[\rho_0(\mathbf{r})] + E^1[\rho_0(\mathbf{r}), \delta\rho(\mathbf{r})] + E^2[\rho_0(\mathbf{r}), (\delta\rho(\mathbf{r}))^2] + E^3[\rho_0(\mathbf{r}), (\delta\rho(\mathbf{r}))^3] \quad (3.13) \end{aligned}$$

^eUnless hybrid functionals like B3LYP are not used.

The zeroth-order term $E^0[\rho_0(\mathbf{r})]$ depends only on the reference density and consists of the DFT double-counting terms. Since this term contains all the core electron effects and solely repulsive contributions, it is also known as the repulsive energy E^{rep} . The E^{rep} term is approximated by a sum of two-body potentials, which are fitted to reproduce DFT or empirical data. These pair potentials V_{ab}^{rep} play a crucial role in all covalent-bonding situations and they are mostly fitted to atomization energies, bond lengths and vibrational frequencies.

$$E^0[\rho_0(\mathbf{r})] \approx E^{\text{rep}} = \frac{1}{2} \sum_{ab} V_{ab}^{\text{rep}} \quad (3.14)$$

The original, non-selfconsistent DFTB1 model^{75,76} is obtained by truncating the Taylor expansion after the first-order term $E^1[\rho_0(\mathbf{r}), \delta\rho(\mathbf{r})]$. This term is also known as the band-structure energy and contains the DFTB Hamiltonian matrix elements. As already mentioned, DFTB exclusively treats valence electrons, and the Kohn-Sham orbitals are expanded in a minimal basis $\{\mu\}$ leading to the energy:

$$E^1[\rho_0(\mathbf{r}), \delta\rho(\mathbf{r})] = \sum_i^{\text{occ}} \langle \psi_i | \hat{H}^0 | \psi_i \rangle = \sum_i^{\text{occ}} \sum_{\mu} \sum_{\nu} c_{\mu}^i c_{\nu}^i H_{\mu\nu}^0 = \sum_i^{\text{occ}} \epsilon_i \quad (3.15)$$

The following approximations are applied for the calculation of the band-structure energy:

- Only valence electrons are considered and each atomic orbital (AO) is represented by only one Slater-type basis function ϕ_{μ} (minimal basis set approach). AOs are well suited for the description of free atoms but they are too diffuse to describe the bonding situation in solids or molecules. In order to emulate this situation and to “contract” the AOs, Eschrig¹¹³ suggested to add an additional harmonic potential to the atomic Kohn-Sham equations:

$$\left[-\frac{1}{2} \nabla^2 + V^{\text{eff}}[\rho_{\text{atom}}] + \left(\frac{\mathbf{r}}{\mathbf{r}^0} \right)^2 \right] \phi_{\mu} = \epsilon_{\mu} \phi_{\mu} \quad (3.16)$$

where $V^{\text{eff}}[\rho_{\text{atom}}]$ contains the electron density of the neutral atom and \mathbf{r}^0 is approximately twice the atom covalent radius.

- The diagonal Hamiltonian matrix elements $H_{\mu\mu}$ are calculated (equation 3.16) by an atomic DFT calculation using the PBE exchange-correlation functional. On the other hand, a two-center approximation is applied for the off-diagonal Hamiltonian matrix elements:

$$H_{\mu\nu}^0 = \left\langle \mu \left| \hat{H}^0 \right| \nu \right\rangle = \left\langle \mu \left| \hat{H}^0[\rho_a(\mathbf{r}) + \rho_b(\mathbf{r})] \right| \nu \right\rangle \quad (3.17)$$

where the orbital μ is located on atom a and orbital ν is located on atom b . As a consequence, crystal-field and three-center terms are neglected. Furthermore, $H_{\mu\nu}$ and $S_{\mu\nu}$ depend only on the distance and orientation of two atoms and are tabulated for all atom pairs at various distances. In contrast to a DFT calculation, these tabulated integrals are read in during a running DFTB calculation which contributes to the efficiency of DFTB.

The total energy of the DFTB1 model then reads:

$$E^{\text{DFTB1}} = \frac{1}{2} \sum_{ab} V_{ab}^{\text{rep}} + \sum_i^{\text{occ}} \epsilon_i \quad (3.18)$$

DFTB1 yields reasonable results for systems where the charges are uniformly distributed like in solids or apolar molecules like hydrocarbons. Furthermore, this model is also able to treat systems which exhibit a complete charge transfer like NaCl⁷⁹. However, for molecular systems where charge is flowing between the atoms the model breaks down. In order to describe these systems, one has to consider these fluctuations in density. This is achieved by adding the second-order term $E^2[\rho_0(r), (\delta\rho(r))^2]$ of the Taylor expansion, leading to the DFTB2/SCC-DFTB⁷⁷ model.

The basis of DFTB2 is the approximation of the density fluctuations $\delta\rho(\mathbf{r})$ as a superposition of atom-centered contributions:

$$\delta\rho(\mathbf{r}) = \sum_a \delta\rho_a(\mathbf{r}) \quad (3.19)$$

Furthermore, a monopole approximation is applied for the atomic contributions:

$$\delta\rho_a(\mathbf{r}) \approx \Delta q_a F_a^{00} Y^{00} \quad (3.20)$$

This means the charge Δq_a is centered at the nucleus position r_a and has a spherical shape like a 1s orbital. On basis of the monopole approximation, the second-order integral is approximated by an analytical function $\gamma_{ab}(r_{ab})$, which interpolates^f between two cases:

- When the distance between the two atoms a and b is large ($r_{ab} \rightarrow \infty$), then the exchange-correlation term vanishes and the integral/ γ_{ab} function describes only the Coulomb interaction between the two fluctuating charges Δq_a and Δq_b .

$$\gamma_{ab}(R_{ab} \rightarrow \infty) = \frac{\Delta q_a \Delta q_b}{r_{ab}} \quad (3.21)$$

- For vanishing interatomic distances ($a = b$ and $r_{ab} \rightarrow 0$) the Coulomb term in the second-order integral vanishes and the integral/ $\gamma_{ab}(r_{ab})$ function describes the electron-electron interaction on atom a :

$$\gamma_{aa}(r_{ab} \rightarrow 0) = \frac{\delta^2 E_a}{\delta^2 q_a} = U_a \quad (3.22)$$

where U_a is known as the Hubbard parameter of atom a , which is related to the chemical hardness η and hence also to the size of the atom by: $U_a = 2\eta_a$. U_a reflects the energy change of an atom with respect to the change of charge, and can be calculated as the difference of the ionization potential and the electron affinity of the isolated atom a . Finally, the DFTB2/SCC-DFTB total energy is given by:

$$E^{\text{DFTB2}} = \frac{1}{2} \sum_{ab} V_{ab}^{\text{rep}} + \sum_i^{\text{occ}} \epsilon_i + \underbrace{\frac{1}{2} \sum_{ab} \Delta q_a \Delta q_b \gamma_{ab}}_{E^2[\rho_0(r), (\delta\rho(r))^2]} \quad (3.23)$$

^fBased on a functional form that was proposed by Klopman and Ohno^{114,115}.

With inclusion of the second-order term, the DFTB formalism has become useful for the description of polar, organic systems and biomolecules. Nevertheless, the monopole approximation still implies three drawbacks:

- The charge density is restricted to a spatial shape, and effects that are based on multipole moments (e.g. interactions of lone pairs) are not treated.
- The assumption that the atom size correlates with the Hubbard parameter holds for most of the main group elements but not for hydrogen. For example, this leads to a bad description of hydrogen-bonded systems. In order to correct for this anomaly a modified γ_{ab}^h function was introduced^{78,79}. This correction is always used when atom a and/or b is hydrogen.
- The Hubbard parameter and hence the atom size is invariant with respect to the charge state of the atom. As a consequence, this impairs the description of charged systems.

A solution to the latter problem is provided by the inclusion of the third-order term $E^3[\rho_0(r), (\delta\rho(r))^3]$ to the DFTB formalism and the introduction of Hubbard derivatives U_a^d with respect to atomic charge. These derivatives are precalculated⁸. The Γ_{ab} function results as the charge-derivative of the Hubbard parameter U_a^d and the DFTB3⁷⁸⁻⁸¹ total energy reads:

$$E^{\text{DFTB3}} = \frac{1}{2} \sum_{ab} V_{ab}^{\text{rep}} + \sum_i^{\text{occ}} \epsilon_i + \frac{1}{2} \sum_{ab} \Delta q_a \Delta q_b \gamma_{ab} + \underbrace{\frac{1}{3} \sum_a \sum_b \Delta q_a^2 \Delta q_b \Gamma_{ab}}_{E^3[\rho_0(r), (\delta\rho(r))^3]} \quad (3.24)$$

The application of the variational principle to the DFTB3 total energy expression with respect to the MO coefficients gives the Kohn-Sham equations. The matrix elements of the Hamiltonian $H_{\mu\nu}$ exhibit a dependence on the charges q , which are given by the MO coefficients. Accordingly, the DFTB Kohn-Sham equations have to be solved in an SCC approach.

In comparison to DFT methods, DFTB runs approximately three orders of magnitude faster and the calculation of ten times larger QM systems (hundreds of atoms) or the prediction of 1000 times longer MD simulations (over nanoseconds) become feasible with DFTB.

⁸Based on a DFT calculation or fitted.

3.3 Treatment of Non-Covalent Interactions in DFTB

As already mentioned in the introduction, a proper treatment of non-covalent interactions like short-range Pauli repulsion and attractive dispersion is indispensable for DFTB in order to achieve chemical accuracy (< 1.0 kcal/mol). Since also *ab initio* methods struggle or are incapable of the description of these interactions, the development of corrections for these interactions has become a big business in quantum chemistry. In order to avoid computationally expensive calculations, these corrections are usually of semiclassical nature and depend rather on atom coordinates than on the electronic structure such that they can simply be added to the quantum-chemical total energy. Thus, the total energy of a DFTB3 model corrected for dispersion and Pauli repulsion effects reads:

$$E_{\text{corr}}^{\text{DFTB3}} = E^{\text{DFTB3}} + E_{\text{disp}} + E_{\text{rep}} \quad (3.25)$$

Regarding dispersion correction, the D3 model is state of the art and was also used in this work. This semiclassical correction is widely used in quantum chemistry and available for most DFT- and semiempirical methods. Contrary to this, the Delta-Pauli model, which should correct for the underestimated Pauli repulsion in DFTB was particularly developed for this method by Maximilian Kubillus⁹⁷. The two following sections will outline how both corrections are modeled.

3.3.1 The D3 Dispersion Correction

From second order perturbation theory it is known that the London dispersion energy between two atoms a and b at large distances r_{ab} is given by:

$$E_{\text{London}} = -\frac{3}{2} \frac{I_a I_b}{I_a + I_b} \alpha_a \alpha_b \frac{1}{r_{ab}^6} \quad (3.26)$$

with I as the ionization potentials and α as the polarizabilities of the respective atoms. For short distances, this attractive interaction vanishes and is superseded by the Pauli repulsion. Dispersion models try to interpolate between both extreme cases. Consequently, the attractive dispersion must be damped at short interatomic distances in order to avoid overbinding in the covalent region. Modern dispersion models hence express the dispersion energy on the basis of r_{ab}^n -dependent (where $n < -6$) terms in combination with a damping function. Furthermore, multipole interactions are unified in empirical parameters C_n^{ab} .

Currently, the D3 model by Grimme et al.⁸⁹ in combination with the Becke-Johnson damping (BJ)^{116–119} is the most popular dispersion correction and was also used in this work. In this model, the atom pairwise dispersion energy is described through a combination of r^{-6}/r^{-8} -dependent terms:

$$E_{\text{disp}}^{\text{D3-BJ}} = -\frac{1}{2} \sum_{ab} s_6 \frac{C_6^{ab}}{r_{ab}^6 + [f_{\text{BJ}}(r_{ab}^0)]^6} + s_8 \frac{C_8^{ab}}{r_{ab}^8 + [f_{\text{BJ}}(r_{ab}^0)]^8} \quad (3.27)$$

with

$$f_{\text{BJ}}(r_{ab}^0) = a_1 r_{ab}^0 + a_2 \quad (3.28)$$

The C_6^{ab} term covers instantaneous dipole-induced dipole interactions, whereas the C_8^{ab} term covers contributions from dipole-quadrupole interactions. s_n are the corresponding scaling factors for these multipolar contributions. The s_6 term dominates the long-range region and is mostly set to unity^h in order to ensure for a correct asymptotic decay of the dispersion energy. Contrary to this, the s_8 scaling factor dominates the short- and medium distances r_{ab} and needs to be fitted for each DFT functional, semiempirical method etc. $f_{BJ}(r_{ab}^0)$ is the BJ damping function, which ensures for the damping of the dispersion energy towards a finite value at short distances. The damping function contains the cutoff radii r_{ab}^0 , which are defined over the multipolar interaction coefficients as: $r_{ab}^0 = \sqrt{\frac{C_8^{ab}}{C_6^{ab}}}$.

3.3.2 The Delta-Pauli Correction

Progress has already been made to describe the short-range Pauli repulsion empirically. Some attempts started with modeling this repulsive interaction on basis of the Born-Mayer functional form^{120,121}:

$$V_{ab}^{\text{rep}}(r_{ab}) = A_{ab} \cdot e^{-B_{ab} \cdot r_{ab}} \quad (3.29)$$

where A_{ab} resembles an overlap proportionality factor of the interacting atom pair ab , and B_{ab} is an atom pair specific exponent. Later on, it will be shown that the Delta-Pauli model, which should correct for the underestimated short-range repulsion is also related to the Born-Mayer model.

In Delta-Pauli the correction energy is modeled as a distance- and charge-dependent repulsive pair-potential:

$$E_{DP} = \frac{1}{2} \sum_{a,b} q_a^0 \cdot q_b^0 \cdot f^{\text{damp}}(r_{ab}) \cdot [1 - \sigma_{ab}(r_{ab})] \cdot V_{ab}^{\text{rep}}(r_{ab}) \quad (3.30)$$

Beside the constant, initial atomic charges q_a^0, q_b^0 this energy expression depends on a damping function f^{damp} , a screening function σ_{ab} and a repulsive potential V_{ab}^{rep} .

Damping Function f_{ab}^{damp}

The damping function is used to switch from the attractive, D3-controlled region ($f_{ab}^{\text{damp}} = 0$) into the repulsive, Delta-Pauli controlled region ($f_{ab}^{\text{damp}} = 1$). A cubic polynomial is applied with the form:

$$f^{\text{damp}}(r_{ab}) = \sum_{i=0}^3 a_i \cdot r_{ab}^i \quad (3.31)$$

For each atom pair, f_{ab}^{damp} is centered around a specific D3 cutoff radius r_{ab}^0 and switches within a specific range from 0 to 1. The width of this switching range is controlled by a predefined parameter called β . In the current Delta-Pauli implementation this parameter is set to $0.1 a_0^i$. In order to obtain the coefficients a_i in equation 3.31 the following linear

^hAlso in DFTB.

ⁱBohr radius.

equation system has to be solved:

$$\begin{aligned} f^{\text{damp}}(r_{ab}^0) &= 0.5 \\ f^{\text{damp}}((1.0 - 0.5\beta) \cdot r_{ab}^0) &= 0.0 \\ f^{\text{damp}}((1.0 + 0.5\beta) \cdot r_{ab}^0) &= 1.0 \end{aligned} \quad (3.32)$$

The behavior of the cubic damping function for the C-H atom pair is shown in Figure 3.1, A.

Repulsive Potential V_{ab}^{rep}

The Delta-Pauli repulsive potential is shown in equation 3.33 and is related to the Born-Mayer equation:

$$V_{ab}^{\text{rep}}(r_{ab}) = s_a \cdot s_b \cdot [e^{-\xi U_a^p(r_{ab}-d)} + e^{-\xi U_b^p(r_{ab}-d)}] \quad (3.33)$$

In the Delta-Pauli framework, the Born-Mayer prefactor A_{ab} is modeled as a product of atom-specific proportionality parameters to describe the electronic density overlap S_{ab} between atom a and b , and the exponent B_{ab} depends on the atomic charges of atom a and b represented by their Hubbard parameters U . More precisely, diffusive Pauli-Hubbards are used that are derived by the standard Hubbard parameters damped by a universal factor α^j :

$$U_x^p = \alpha \cdot U_x \quad (3.34)$$

Figure 3.1, B depicts how the Delta-Pauli repulsive potential reacts upon change of S_{ab} and ξ .

The last remaining universal parameters in equation 3.33 is d . This parameter controls the interatomic distance shift of the potential.

Screening Function σ_{ab}

The screening function represents the main ingredient of the Delta-Pauli theory, since it makes Delta-Pauli geometry-sensitive and “decides” which interactions are corrected by it and which are not. For example, consider the *cis*- and *trans*-conformation of n-butane in Figure 3.2B/C. The rotational barrier along the dihedral angle $a - c_1 - c_2 - b$ is underestimated in DFTB3 due to missing short-range repulsion. Especially the *cis*-conformation, where atoms a and b come very close to each other suffers from the missing repulsion, so that the barrier height of DFTB3 has the largest error for the *cis*-conformation. Hence, the Delta-Pauli correction should especially improve the representation of systems like *cis*-n-butane. On the other hand, there should be no contribution for systems like the *trans*-n-butane where the barrier height is represented correctly with DFTB3 (atom a and b are far apart from each other) and no further repulsion is needed. The overlap-dependent screening function σ_{ab} accounts for this geometry specificity and is able to discriminate between such

^jWe also tested repulsive potentials that either used atom pair dependent overlap-proportionality parameters s_{ab} or atom-specific Hubbard-scaling factors α_x but equation 3.33 yielded the best performance.

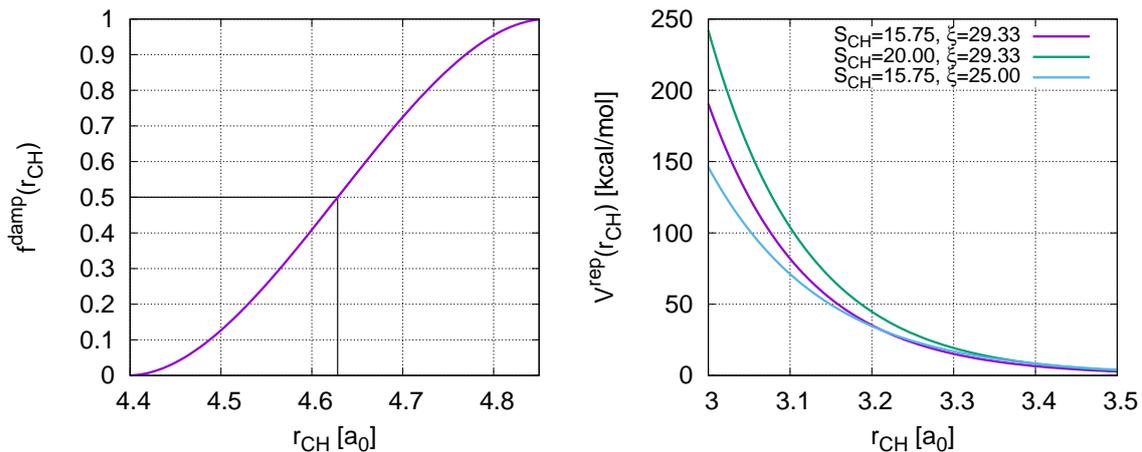


Figure 3.1: Functional behavior of the Delta-Pauli damping function f_{ab}^{damp} (A) and Delta-Pauli repulsive potential V_{ab}^{rep} (B) for the CH atom pair. A: $f^{\text{damp}}(r_{CH})$ switches within a specific range around $r_{CH}^0 = 4.63$ from 0 to 1 and consequently activates/deactivates Delta-Pauli. B: A change of the electronic density overlap product $S_{CH} = s_C \cdot s_H$ increases/decreases V_{CH}^{rep} at short-distances, whereas a change of the ξ parameter affects the steepness of the function.

cases through the introduction of a virtual third atom density c . This atom density in- or decreases the effect of the Pauli repulsion correction between atoms a and b :

$$\sigma_{ab}(r_{ac}, r_{bc}, r_{ac}^*, r_{bc}^*) = \sum_c^N \left[\frac{(r_{cc}^0)^2}{r_{aa}^0 \cdot r_{bb}^0} \right]^\zeta \cdot \frac{S_{ac}(r_{ac}) S_{bc}(r_{bc})}{S_{ac}^*(r_{ac}^*) S_{bc}^*(r_{bc}^*)} \quad (3.35)$$

where N is the number of atoms. The ζ -dependent prefactor is based on a hard spheres model with the atom covalent radii r_{xx}^0 as components. ζ is a universal parameter that controls how fast σ_{ab} decays towards 0. The latter quotient depends on the overlap of the atom densities a and b with c . The factors S_{xc} depend on the distance r_{xc} , while the $*$ is the projection of the atom density c on the connecting line between atom a and b (see Figure 3.2A). To model the overlap S_{xc} we use the short-range part of the DFTB3- γ^H function. We can get a better understanding for σ_{ab} when we take the n-butane system as example: The third density c would then account for the presence of the atoms c_1 and c_2 . The *cis*- and *trans*-conformation represent the two possible extreme cases of the σ_{ab} -function:

- *cis*-n-butane: The third density c does not lie on the connecting line between atoms a and b (since c_1 and c_2 do neither), so that the overlap-dependent quotient and thus σ_{ab} would be close to 0. This leads to the addition of the full, undamped Delta-Pauli energy according to equation 3.30.
- *trans*-n-butane: The connecting line between atom a and b is interrupted by atoms c_1 and c_2 . This leads to $\sigma_{ab} \approx 1$, since the overlap-dependent quotient is also close to 1. This conformation would not be affected by the correction since the Delta-Pauli energy would be fully damped (cf. equation 3.30).

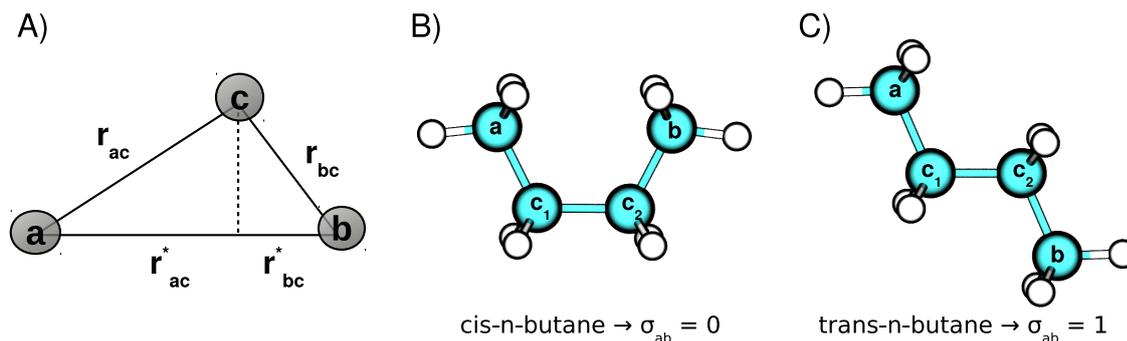


Figure 3.2: A: The screening behavior of the Delta Pauli σ_{ab} function. The virtual density C regulates the effect of the Delta Pauli correction applied to the atom pair A and B. The closer C lies to the connecting line between A and B, the more strongly σ_{ab} screens the effect of Delta-Pauli (A). Consequently, the screening function evaluates to $\sigma_{ab} = 0$ for the *cis*-n-butane (B) and to $\sigma_{ab} = 1$ for the *trans*-n-butane (C). See text for further explanation.

Particle Swarm Optimization

In order to apply the Delta-Pauli correction to CHNO-based systems, several parameters need to be derived first. There are several algorithms for solving such multidimensional optimization problems and some of them map the problem down to the minimization of a given fitness/cost function f :

$$f(\vec{x}) : \mathbb{R}^D \rightarrow \mathbb{R} \quad (3.36)$$

where \mathbb{R}^D represents the D-dimensional parameter space and every point in it can be described by a vector of coordinates \vec{x} . The objective of the optimization is to find an optimal solution \vec{x}_{opt} which minimizes f ¹²²:

$$f(\vec{x}_{opt}) \leq f(\vec{x}) \quad \forall \vec{x} \in \mathbb{R}^D \quad (3.37)$$

Some of them like genetic algorithm¹²³, bayesian optimization¹²⁴ or particle swarm optimization (PSO)¹²⁵ have already been used for DFTB parameter development. In this work the Delta-Pauli and D3 dispersion parameters will be also derived via PSO, which was introduced by Kennedy and Eberhardt in 1995¹²⁶ and further developed by Shi¹²⁷. The algorithm is based on a swarm of particles, where each particle represents a possible parameter set $\vec{x} \in \mathbb{R}^D$. At the beginning, the particles are placed randomly in the parameter space. From time to time the particles change their positions semistochastically with the purpose of gradually decreasing the fitness function f . In the course of this, the particles depend on each other: One particle P_x tries to move towards its personal best parameter combination \vec{x}_p while it is also dragged by the particle P_g with the global best parameter combination \vec{x}_g towards its position. If the new position of P_x yields a better parameter set, then this position will become the new \vec{x}_p . The travel of the particles is controlled by four PSO constants that need to be defined beforehand: the total number of particles N , the inertia weight ω , the cognitive acceleration constant ϕ_p and the social acceleration

constant ϕ_g . The inertia weight ω controls the velocity of each particle and hence the local exploration and exploitation. The acceleration constants ϕ_p and ϕ_g determine the tendency to move towards the local (\vec{x}_p) and global (\vec{x}_g) minimum respectively^{128,129}. A PSO step is shown schematically in Figure 3.3.

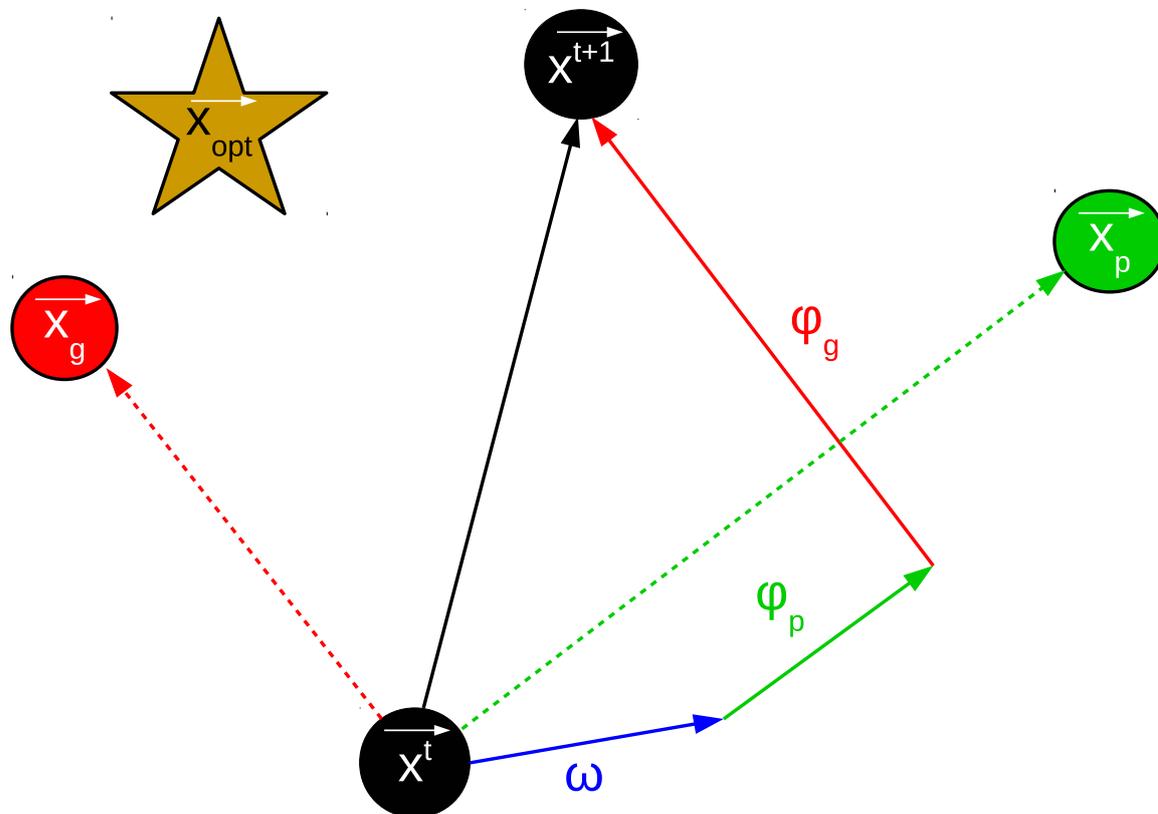


Figure 3.3: Example for a PSO step. The global best position \vec{x}_g is near to the global minimum \vec{x}_{opt} which needs to be found. The particle P_x starts from the position \vec{x}^t . Due to its inertia weight ω it is dragged in an arbitrary direction. The “flight direction” is moreover influenced by the attraction of the local best position \vec{x}_p and global best position \vec{x}_g . In the next PSO step, the particle P_x is initiated at position \vec{x}^{t+1} which is closer to the global minimum.

Molecular Mechanics

Although semi-empirical methods exhibit a significant speedup, they are inappropriate for the description of entire molecules like proteins, membranes etc. because of two reasons: On the one hand, the system size with more than multiple thousands of atoms is simply too expensive to calculate. On the other hand, biomolecular processes of interest like protein folding etc. occur on a too long timescale of up to multiple seconds. Nevertheless, the simulation of these large systems and timescales can be performed with molecular mechanics (MM). In this framework the electronic structure is effectively described by point charges such that the energy of the system can be expressed as a function of the nuclear coordinates \mathbf{R} . Consequently, the molecular system is represented by atom-centered mass points. The interaction of the atoms is treated classically, and is modeled empirically via parameterized potentials.

The atomic parameters in force fields are usually defined through atom types. This means the properties of an atom depend not just on the element but also on its bonding situation and chemical environment. Accordingly, different parameters will be used for an N-H bond in ammonia and an N-H bond in a peptide etc. Only very few force fields are parameterized for the purpose of large transferability in order to yield reasonable results for systems comprising elements of the whole periodic table^a. In this work the CHARMM36 force field^{130,131} was used, which was specially parametrized for proteins.

^ae.g. Universal Force Field (UFF).

4.1 The CHARMM Force Field

The CHARMM force field energy is, like in the most force fields, modeled as a sum of bonded and non-bonded contributions:

$$\begin{aligned}
 E^{\text{CHARMM}} &= E_{\text{bonded}} + E_{\text{non-bonded}} \\
 &= \sum_{\text{bonds}} k_R (b - b_0)^2 + \sum_{\text{angles}} k_\theta (\theta - \theta_0)^2 + \sum_{\text{dihedrals}} k_\phi [1 + \cos(n\phi - \delta)] \\
 &+ \sum_{\text{impropers}} k_\omega (\omega - \omega_0)^2 + \sum_{\text{Urey-Bradley}} k_u (u - u_0)^2 + \sum_{\text{residues}} E_{\text{CMAP}}(\phi, \psi) \\
 &+ \sum_{\text{non-bonded}} \left\{ \underbrace{\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{R_{ij}} \right)^{12} - 2 \left(\frac{\sigma_{ij}}{R_{ij}} \right)^6 \right]}_{\text{van-der-Waals}} + \underbrace{\frac{q_i q_j}{4\pi\epsilon_0 R_{ij}}}_{\text{Coulomb}} \right\} \quad (4.1)
 \end{aligned}$$

The bonded energy term contains, like the name suggests, interactions of atoms that are mediated by bonds. The fluctuation of a chemical bond b between the atoms A and B is approximated via a harmonic potential with the force constant k_b around the equilibrium bond distance b_0 (see Figure 4.1A). Consequently, this model gives reasonable results for small deviations around b_0 but is incapable of the description of a bond dissociation.

Another harmonic approximation is made for the displacement of the bond angle θ defined by the atoms A , B and C (see Figure 4.1B).

A torsional angle is defined via four atoms A , B , C and D (see Figure 4.1C). Since the torsional angle is periodic, also the potential energy has to be modeled in a periodic fashion. In order to achieve this, the CHARMM force field employs a cosine function. The periodicity of the potential is described by the dihedral angle prefactor n (e.g. $n = 3 \equiv 120^\circ$). Furthermore, the height of the torsional barrier is given by the force constant k_ϕ and the phase shift by δ .

Since force fields neglect QM effects, special bonding situations like a planar π -conjugated system need an additional force field term that ensures the planarity. Therefore the improper dihedral term is introduced. The improper dihedral angle ω is defined by atoms A , B , C , D and models the interaction of the atoms A , B , D to the central atom C via a harmonic potential (see Figure 4.1D).

Moreover, the CHARMM force field employs the Urey-Bradley term in order to describe 1,3-interactions that are necessary for the correct description of vibrational spectra. The term involves the 1,3-distance between atom A and C , and is modeled as a harmonic potential.

A force field term, which is unique to the CHARMM force field, is the CMAP¹³² energy. It represents a quantum mechanically derived energy correction for the protein backbone dihedrals ϕ , ψ . As a consequence, the description of peptide conformations is improved.

The last term in equation 4.1 summarizes the contributions of the non-bonded interactions that are mediated through space: van-der-Waals and electrostatic interactions. The former is modeled by the Lennard-Jones potential, which describes the attractive dispersion

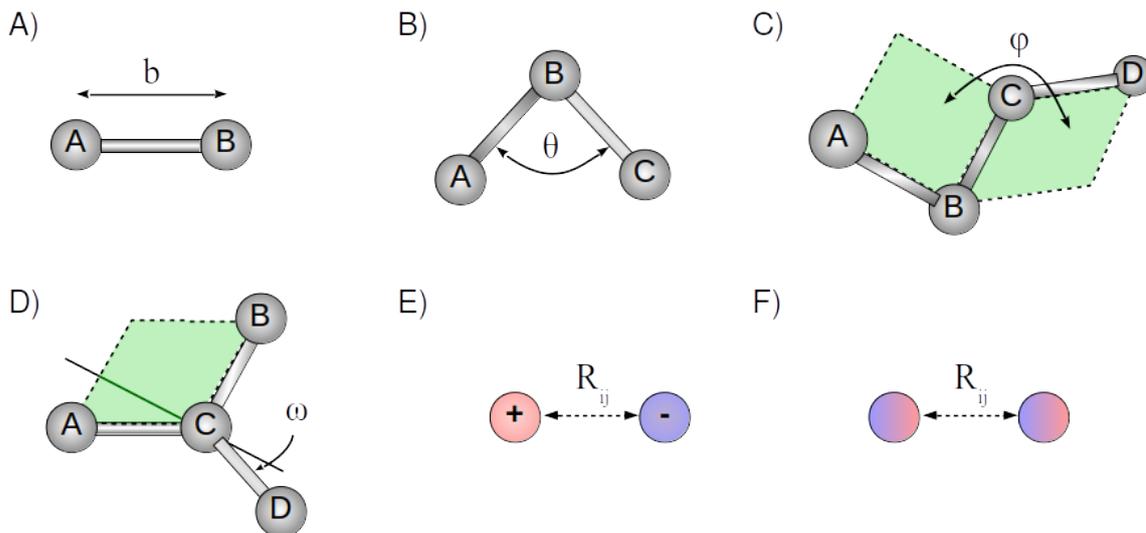


Figure 4.1: Illustration of force field terms. A: Bond stretching. B: Angle bending. C: Torsional bending. D: Improper dihedral bending. E: Electrostatic interaction. F: van-der-Waals interaction.

interaction as well as the Pauli-repulsion of two atoms i and j in dependence of their distance R_{ij} (see Figure 4.1F). σ_{ij} and ϵ_{ij} are empirical parameters, which specify the depth and the location of the potential minimum.

The electrostatic interaction of two point charges q_i and q_j at a distance R_{ij} (see Figure 4.1E) is described via Coulomb's law, where ϵ_0 is the permittivity in vacuo. It is worth mentioning that the computation of the non-covalent interactions is the most time-consuming part in the calculation of the force field energy since these interactions need to be evaluated for all atom pairs. Furthermore, whereas the dispersive interactions decay with a \mathbf{R}^{-6} dependence, the electrostatic interactions converge very slowly ($\propto \mathbf{R}^{-1}$). In order to optimize these calculations, cutoff techniques and more sophisticated methods like the particle mesh Ewald^{133,134} method have been developed.

4.2 Molecular Dynamics

While quantum chemical methods are mostly applied for optimizing characteristic geometries^b of a molecule on its potential energy surface (PES), force field methods are predominantly used in MD simulations. In order to obtain a trajectory of a molecular system consisting of i particles, the Newtonian equations of motion have to be solved:

$$\mathbf{F}_i = m_i \frac{d^2 \mathbf{R}_i}{dt^2} \quad (4.2)$$

The force \mathbf{F}_i provides the acceleration $\frac{d^2 \mathbf{R}_i}{dt^2}$ of the i th atom with mass m_i . In a propagation step the acceleration is combined with the atom position $\mathbf{R}_i(t)$ and velocity $\mathbf{v}_i(t)$ to yield its

^be.g. minima, transition states etc.

new position $\mathbf{R}_i(t + \Delta t)$ and velocity $\mathbf{v}_i(t + \Delta t)$. The force \mathbf{F}_i is evaluated as the negative derivative of the force field energy with respect to the atoms position:

$$\mathbf{F}_i = -\frac{\partial E(\mathbf{R}_i)}{\partial \mathbf{R}_i} \quad (4.3)$$

For a system of i atoms, this leads to i differential equations which can not be solved analytically and instead must be solved numerically using finite difference algorithms. In this work, the leap-frog integrator¹³⁵ was used that evaluates first the velocities at half time steps and then the coordinates at full time steps:

$$\begin{aligned} \mathbf{v}\left(t + \frac{1}{2}\Delta t\right) &= \mathbf{v}\left(t - \frac{1}{2}\Delta t\right) + \mathbf{a}(t)\Delta t \\ \mathbf{R}(t + \Delta t) &= \mathbf{R}(t) + \mathbf{v}\left(t + \frac{1}{2}\Delta t\right)\Delta t \end{aligned} \quad (4.4)$$

The choice of the time step Δt is dictated by the fastest motions of the molecular system, which are usually the vibrations of hydrogen atoms. In order to resolve these motions properly the time step in MD simulations is commonly 1 fs. With constraint algorithms^c like SHAKE¹³⁶, LINCS¹³⁷ etc. the time step can be enlarged to 2 fs, which was also used in this work. The propagation of a molecular system with the equations of motion shown in equation 4.4 will lead to the microcanonical/NVE ensemble. However, for the comparison of a MD simulation with experimental data or for the prediction of experimental values, the sampling has to happen in an ensemble that corresponds to the experimental reality. Since most experiments are conducted under constant temperature and constant pressure conditions, MD simulations are usually designed to generate the NPT/isothermal-isobaric ensemble. Therefore, additional algorithms are needed that ensure for the temperature and pressure coupling. In this work the Nosé-Hoover thermostat¹³⁸⁻¹⁴⁰ was used. This algorithm treats the heat bath as an internal part of the system and introduces it via an additional degree of freedom, which can be interpreted as a friction term.

The barostat algorithms control the pressure of the system over its volume variation. Again, this can be achieved directly through volume scaling or through an additional degree of freedom like in the Parinello-Rahman¹⁴¹ barostat, which was used in this work. Compared to other thermostats and barostats that introduce temperature and pressure by simply scaling atom velocities^d and variation of the box volume^e, the Nosé-Hoover thermostat and Parinello-Rahman barostat do generate a canonical NPT ensemble.

^cApplied to keep the lengths of X-H bonds fixed during the simulation.

^de.g. Berendsen thermostat.

^ee.g. Berendsen barostat.

4.3 Quantum Mechanics/Molecular Mechanics

The strengths and deficiencies of quantum chemical- and force field-based methods have been described in the two previous sections. However, for the study of chemical reactions, electron transfer processes, photochemical processes etc. these methods are either inefficient or incapable. A solution to this problem yields the combination of both methods and their strengths, and results in the hybrid QM/MM approach. The idea behind this method is based on the partition of the whole system in subsystems, which are then treated with different level of theory. For example, the reaction taking place in an enzyme-substrate-complex is mostly localized in an active site. A QM description is required for this region because of the rearrangement in the electronic structure. For the remaining, non-reactive and much bigger part of the protein or system (solvent etc.), a classical force field-based description is sufficient. As a consequence, the Hamilton operator of the whole system is split in contributions of the QM and MM system and their interactions (additive QM/MM scheme):

$$\hat{H}_{\text{total}} = \hat{H}_{\text{QM}} + \hat{H}_{\text{MM}} + \hat{H}_{\text{QM/MM}} \quad (4.5)$$

While the calculations of the pure MM and QM subsystems are straightforward, the description of the QM/MM interaction is more difficult. Especially the treatment of the electrostatic interactions between the QM and MM region is important and technically the most involved one. There exist three main approaches to deal with it:

- Mechanical embedding: The electrostatic interaction between the regions is modeled like in force fields as an electrostatic interaction of point charges. This is computationally efficient and technically straightforward. However, the biggest problem is that the QM region does not get polarized by the MM region.
- Electrostatic embedding: In this scheme, the MM point charges enter the QM Hamiltonian as one-electron terms. As a consequence, the QM region adapts to changes in the MM region and is polarized by it. Since the electrostatic interaction is treated at the QM level, this embedding scheme is in some cases computationally more expensive. This embedding scheme is the most popular one and was also used in this work.
- Polarization embedding: This embedding scheme also accounts for the polarization of the MM region through the QM region (and back). However, the computational cost and the absence of an established polarizable force field makes this scheme an difficult alternative.

For the other interactions (bonded and van-der-Waals) between the QM and MM region the force field representation is used.

The setup of a QM/MM simulation implies the presence of a QM/MM border. In most cases, this border passes through a covalent bond such that the QM region is not a chemically meaningful structure. There are various approaches available in order to saturate the QM region. A conceptually simple and hence prominent approach is the introduction

of link atoms, which was also used in this work. In this scheme, the valence of the QM atom at the QM/MM boundary is saturated by a virtual hydrogen atom (the link atom) and only reinserted in the QM region. As a consequence, in the QM calculation forces act on the link atoms which must be redirected to the neighboring atoms. Since the electron density of the link atom is spatially very close to the MM point charges, the QM region may get overpolarized. The divided frontier charge (DIV) scheme¹⁴² counteracts this overpolarization due to the manual charge spreading of the MM boundary atom.

Enhanced Sampling and Proton Transfer Reactions

The fundamental aim of MD simulations is the correct sampling of the $6N$ -dimensional phase-space density $\rho(\mathbf{r}, \mathbf{p})$, consisting of the coordinates \mathbf{r} and momenta \mathbf{p} of the atoms, in order to predict at given conditions interesting properties of the system like for example free energies. The justification for this represents the ergodic theory^{143–145}, which says that the time average $\langle A \rangle_t$ which is provided by a simulation, is equal to the ensemble average $\langle A \rangle_e$ of a property A , which is measured with most experimental techniques:

$$\langle A \rangle_t = \langle A \rangle_e \quad (5.1)$$

This means, thermodynamic ensemble properties can be obtained through the trajectory of a long enough simulation, which passes through all the phase space points with non-negligible densities (ergodic simulation) such that equation 5.1 becomes:

$$\lim_{t \rightarrow \infty} \frac{1}{t} \int_{t_0}^{t_0+t} A(t) dt = \iint A(\mathbf{r}, \mathbf{p}) \cdot \frac{1}{Q} \cdot \exp \left[\frac{-E(\mathbf{r}, \mathbf{p})}{k_B T} \right] d\mathbf{r} d\mathbf{p} \quad (5.2)$$

with t as the total simulation time, E as the systems total energy, k_B as Boltzmann's constant, T as the temperature and Q as the system partition function.

The ergodic hypothesis offers a great potential for MD simulations, however it poses at the same time the biggest problem. For example, consider a movement of a flexible amino acid in a protein which can be represented by a reaction coordinate/collective variable $\zeta(\mathbf{r})$ like, for example, a distance. The energy profile for this process with respect to $\zeta(\mathbf{r})$, can then be mapped down like schematically shown in Figure 5.1. The energy profile exhibits, in this picture, two stable conformations/minima for the amino acid. Structural rearrangements of amino acids usually lie on a time scale of multiple nanoseconds¹⁴⁶, which is feasible to simulate. However, MD simulations at room temperature hardly overcome barriers greater than several $k_B T$, since the probability of “hitting” the correct degrees of freedom which lead to a barrier crossing is very low. With respect to Figure 5.1 this means the simulation is maybe started in conformation ① but during simulation never overcomes the barrier which leads to the representative global minimum conformation ②.

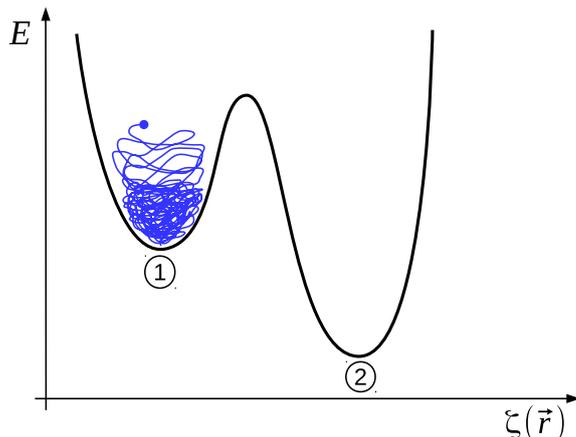


Figure 5.1: Illustration of the “sampling problem” based on a model Gibbs (free) energy surface. A simulation is initiated in the local minimum ①. In the course of simulation the system never overcomes the barrier leading to the global conformation ②.

As a result, one predicts the wrong representative conformation and - even worse - one may deduce wrong reaction mechanisms. Another difficulty that this so-called “sampling problem” leads to, is the prediction of interesting thermodynamic properties like Gibbs (G) or Helmholtz (F) free energies. For example, one is interested in the free energy barrier ΔG^\ddagger connecting basins ① and ②. If one wants to give an estimate of this ensemble property $\langle A \rangle_e$ the energy landscape connecting both basins has to be sampled in order to get values for $E(\mathbf{r}, \mathbf{p})$ (cf. equation 5.2). This is unproblematic for the minima basins ① and ②, since they are well sampled during simulation. However, the higher lying conformations either do not even get sampled (like in Figure 5.1) or are undersampled since the simulation time is too short. As a result, in the first case, one simply can not give an estimate for ΔG^\ddagger , whereas in the second case, the estimate will probably contain a huge error. This is because the most important data points for the ΔG^\ddagger prediction are the energetically higher lying $\zeta(\mathbf{r})$ conformations since these will contribute significantly^a. Consequently, in the field of MD simulations, there is a need for techniques that are able to bypass the sampling problem in order to predict properties of a molecular system correctly and find its representative conformations. In the past few decades various of these so-called enhanced sampling techniques have been developed. The ones which were used in this work are presented in the following. Some of them like Metadynamics and Umbrella Sampling need a predefined collective variable/reaction coordinate, which provides a reasonable description of the process. The sampling is then enhanced in this collective variable space. The other ones like Simulated Annealing and Replica Exchange increase the sampling in the whole phase-space.

Since this work focuses on the study of PT reactions, a proper reaction coordinate for this process is needed and presented at the end of this chapter.

^aFree energies F are given by $F = k_B T \ln \left\langle \exp \left[\frac{E(\mathbf{r}, \mathbf{p})}{k_B T} \right] \right\rangle - \ln c$. Thus, large $E(\mathbf{r}, \mathbf{p})$ values enter an exponential function and will contribute significantly to free energy estimates.

5.1 Simulated Annealing

Simulated annealing¹⁴⁷ is from the technical perspective probably one of the most straightforward approaches to overcome barriers. With respect to equation 5.2 the probability to visit a phase-space point $P(\mathbf{r}, \mathbf{p})$ with a specific energy E shows the following proportionality:

$$P(\mathbf{r}, \mathbf{p}) \propto \exp \left[\frac{-E(\mathbf{r}, \mathbf{p})}{k_B T} \right] \quad (5.3)$$

This dependence can be exploited: In order to increase the probability for visiting higher energetical phase space points and hence overcome barriers, simply the temperature needs to be increased. Figure 5.2 depicts schematically a Simulated Annealing protocol for the model energy profile which was already depicted in Figure 5.1. The Simulated Annealing run starts with our molecular system trapped in the local minimum ① at temperature T_{low} (see Figure 5.2A). Over the course of simulation, the temperature is linearly increased to a target temperature T_{high} which enables the system to overcome the barrier such that it samples both minima (see Figure 5.2B). Moreover, the probability to find the system in the global minimum ② is larger than for finding it in basin ①. Thus, there is a bigger chance to end up in the global minimum after cooling down the system back to T_{low} (see Figure 5.2C). However, there is no guarantee that the simulated annealing finds the true global minimum. Moreover, it is not quite clear, how to estimate the barrier height (e.g. ΔG^\ddagger) for the process ① \rightarrow ② at T_{low} .

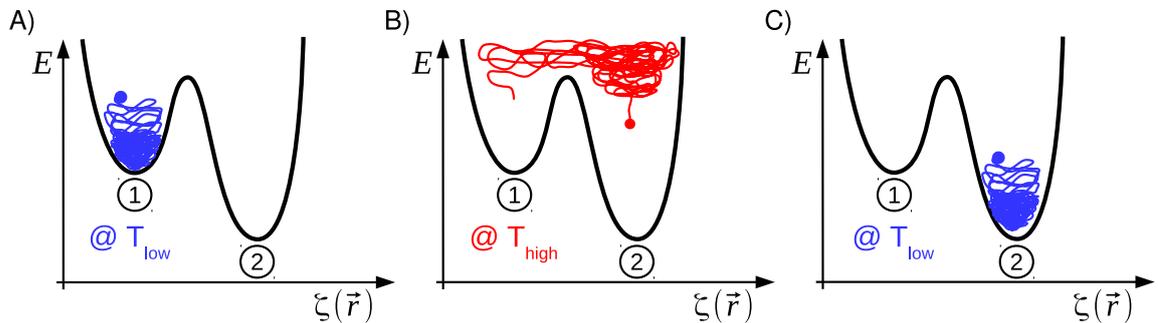


Figure 5.2: Enhanced phase space sampling with Simulated Annealing. In the course of the simulation, the temperature is increased (A), which enables the system to escape the local minimum ① and sample a larger region in the phase space (B). At the end, the system is cooled down to the initial system temperature and probably is trapped in the global minimum ② (C).

5.2 (Hamiltonian) Replica-Exchange

Another possibility to overcome barriers and to predict their heights is provided by Replica Exchange MD (REMD) simulations. Like in simulated annealing, also this method works by distorting the phase-space probability shown in equation 5.3. In this approach, several copies (replicas) of the molecular system are simulated in parallel. The difference between these replicas is the accessible phase space. In the original implementation of REMD¹⁴⁸ this was ensured by using different temperatures for each replica^b. As a consequence, with respect to Figure 5.3 the replica which samples at the reference temperature (lowest temperature) hardly overcomes the barrier and the conformational exploration is restricted to basin ① (see Figure 5.3A). Importantly, the replicas at higher temperatures are able to explore larger parts of the conformational space (see Figure 5.3B+C). The key step is the exchange attempt of the positions \mathbf{r} and momenta \mathbf{p} of two replicas after a predefined time τ (see Figure 5.3D). Thereby, only replicas which are similar in their potential energies^c and therewith fulfill a specific acceptance ratio^d undergo an exchange such that in the course of the simulation the replicas exchange often. As a consequence, the replica at the temperature of interest (reference) still generates the correct ensemble and is furthermore provided with “conformational information” from higher tempered and hence better sampled replicas.

In this work the Hamiltonian Replica Exchange (HREX)^{149–151} method is used. This REMD variant scales, instead of temperatures (denominator in equation 5.3), the Hamiltonians (numerator in equation 5.3) between the different replicas. More precisely in the HREX method Coulomb, Lennard-Jones and proper dihedral terms are scaled by a factor λ since these interactions are significantly contributing to energy barriers. For the unscaled reference system a λ factor of 1.0 is used, whereas the replicas where sampling is increased exhibit a λ factor smaller than 1.0. In contrast to temperature, energy is an extensive property which allows to apply the scaling to selected parts of the system - the so-called “hot regions”. Like Simulated Annealing, also REMD increases the sampling in the whole phase space, but in contrast with REMD it is possible to predict free energy differences^e.

^bThis approach is also known as parallel tempering.

^cThese are mostly neighboring replicas.

^de.g. based on the Metropolis criterion.

^eFor example, the Gibbs free energy difference between the two basins ① and ② of the model energy profile

shown in Figure 5.1 is given by their probabilities of presence: $\Delta G = k_B T \cdot \ln \left(\frac{P(\textcircled{1})}{P(\textcircled{2})} \right)$

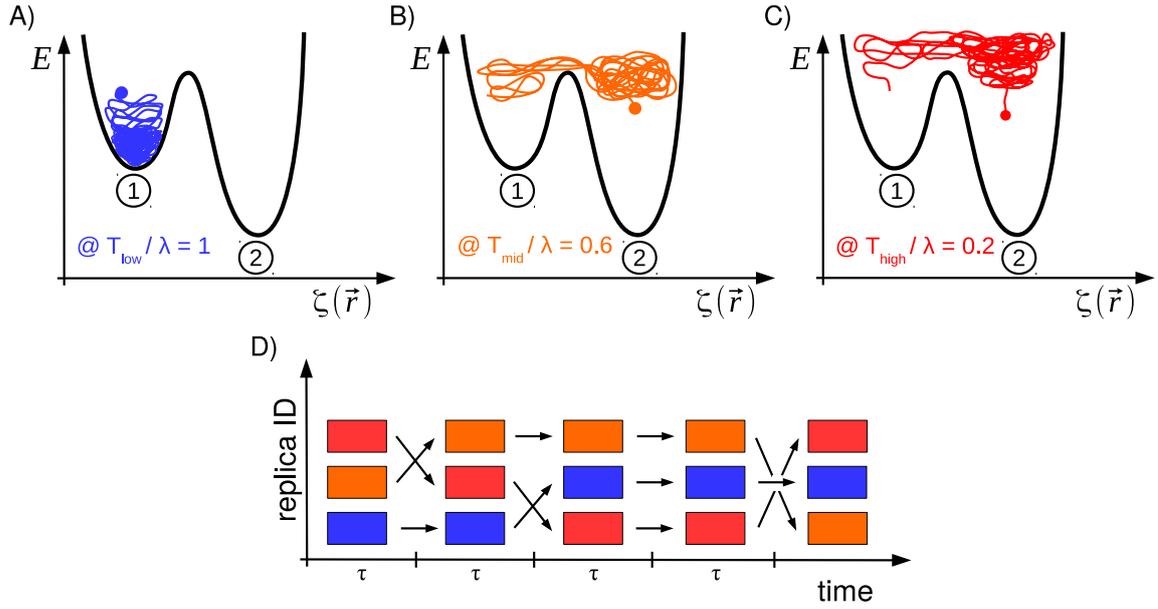


Figure 5.3: Enhanced phase space sampling with Replica Exchange methods. A set of replicas is simulated at a reference temperature/with an unscaled Hamiltonian (A) and at higher temperatures/with downscaled Hamiltonians (B+C). The phase space sampling increases at higher temperatures/lower λ factors such that barriers are overcome and both minimum conformations ①/② are sampled. After a predefined time τ the replicas exchange their coordinates and velocities (D).

5.3 Umbrella Sampling

In Umbrella Sampling simulations, the change in free energy along a predefined reaction coordinate ζ is resolved. To this end, the free energy is defined as a function of this reaction coordinate, often called the PMF¹⁵²:

$$F(\zeta_0) = -k_B T \cdot \ln P(\zeta_0) + \text{const.}$$

$$\text{with } P(\zeta_0) = \frac{\int \delta[\zeta - \zeta_0] \exp\left[\frac{-E(\mathbf{r}, \mathbf{p})}{k_B T}\right] d\mathbf{r} d\mathbf{p}}{\int \exp\left[\frac{-E(\mathbf{r}, \mathbf{p})}{k_B T}\right] d\mathbf{r} d\mathbf{p}} \quad (5.4)$$

with k_B as the Boltzmann constant, T as the temperature and $P(\zeta_0)$ representing the probability to find the system in a small interval $d\zeta$ around ζ_0 . The probability distribution is obtained by integration over the whole phase space except over ζ^f . In MD simulations these phase space integrals are approached by time averages (cf. equation 5.1). This means, the free energy estimate for a system located at ζ_0 can be obtained by simply counting how often the (long enough) MD simulation populates ζ_0 . However, this approach works only if the ζ_0 conformation is well sampled/low in energy (e.g. ① and ② basin in the model energy profile). For ζ -structures which lie significantly higher than $k_B T$ the sampling is unfeasible and the free energy estimate will exhibit a huge error.

^fThe constant is mostly chosen such that the free energy of the most probable distribution is equal to zero.

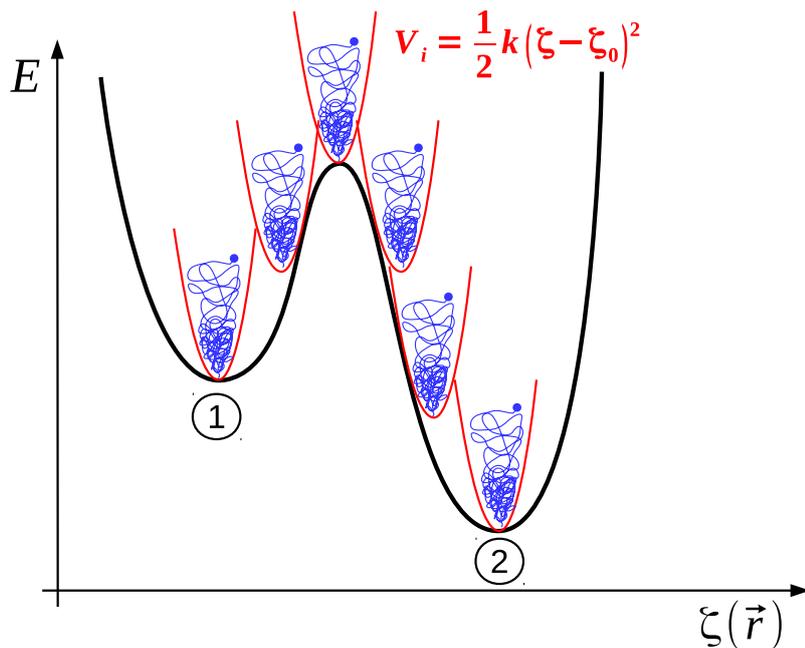


Figure 5.4: Enhanced phase space sampling with Umbrella Sampling. The energy landscape is explicitly defined as a function of the reaction coordinate ζ (PMF). The sampling at specific locations ζ_0 along the predefined reaction pathway can be artificially increased by using a harmonic bias potential which is centered at these positions. In order to resolve the complete reaction pathway multiple simulations – so-called “windows” – are used which distinguish in their location of the harmonic bias potential.

The Umbrella Sampling¹⁵² approach allows for the sampling of these high-energy conformations via the introduction of an additional bias potential $V(\zeta)$, which may be centered at specific ζ positions, in order to increase the sampling in this region. Usually, $V(\zeta)$ is a harmonic potential with the form:

$$V(\zeta) = \frac{1}{2}k(\zeta - \zeta_0)^2 \quad (5.5)$$

For an efficient sampling along the ζ coordinate, the system is replicated into so-called “windows” and each of these is simulated at shifted $V(\zeta)$ positions (see Figure 5.4). Conveniently, the unbiased free energy $F_i(\zeta)$ in the i th window is given by its observed probability P_i^b in the biased simulation minus the additional potential $V_i(\zeta)$:

$$F(\zeta) = -k_B T \ln P_i^b(\zeta) - V_i(\zeta) + \text{const.}_i \quad (5.6)$$

The first term in equation 5.6 is obtained from the biased simulation directly, the second term is just the biasing potential that has to be subtracted and the constant is unknown but independent of ζ . In order to obtain the complete free energy profile along the ζ coordinate, the individual windows have to be combined and the constants const._i need to be calculated, for example with the weighted histogram analysis method (WHAM)¹⁵³. A prerequisite for this is a sufficient overlap of the biased probabilities $P_i^b(\zeta)$ and $P_{i+1}^b(\zeta)$ between neighboring windows.

5.4 Metadynamics

Another approach which allows the exploration and reconstruction of a free energy surface is Metadynamics¹⁵⁴. The key ingredient of this method is the use of an external, history-dependent bias potential $V(\zeta, t)$ which is added to the Hamiltonian of the system and only acts on a predefined set of collective variables $\zeta_i(\mathbf{r})$. In practice, the bias potential is constructed as a sum of Gaussians which are deposited along the trajectory. As a result, conformations that have already been sampled become penalized energetically and the system is pushed to other regions of conformational space. Consequently, due to the “filling up” of the conformational basins, the system discovers new energy minima. The bias potential $V(\zeta, t)$ at a specific time t is given by:

$$V(\zeta, t) = \int_0^t dt' \omega \exp \left[- \sum_{i=1}^d \frac{(\zeta_i(\mathbf{r}) - \zeta_i(\mathbf{r}(t')))^2}{2\sigma_i^2} \right] \quad (5.7)$$

with σ_i as the Gaussian width and ω as the Gaussian deposition rate.

In the well-tempered Metadynamics¹⁵⁵ implementation the Gaussian height decreases with simulation time such that ω reads:

$$\omega = \omega_0 \tau \exp \left[- \frac{V(\zeta, t)}{k_B T} \right] \quad (5.8)$$

with ω_0 as the initial Gaussian height and τ as their deposition frequency. The parameters σ_i , ω_0 and τ have to be defined a priori and they control the resolution and convergence rate of the free energy profile¹⁵⁶.

In order to further understand the effect of $V(\zeta, t)$ consider again the model free energy profile in Figure 5.5. The well-tempered Metadynamics simulation starts in the local minimum ①. Gaussians are deposited in the ζ space in regular intervals, filling up the free energy profile. After a certain time t_2 the system is able to escape from the local minimum via the lowest free energy saddle point and falls into the global minimum ② where the procedure continues. Consequently, in the long-time limit the bias potential smoothly converges to the underlying free energy profile:

$$V(\zeta, t \rightarrow \infty) = - \frac{\Delta T}{T + \Delta T} F(\zeta) + C \quad (5.9)$$

where $F(\zeta)$ is the free energy profile in the ζ space, T is the simulation temperature and C is an additive constant. The parameter ΔT is a priori defined over the bias factor:

$$\gamma = \frac{\Delta T + T}{T} \quad (5.10)$$

This parameter allows to define the maximum temperature range $T + \Delta T$ in which the collective variables are sampled. A bias factor of $\gamma = 1$ ($\Delta T = 0$) corresponds to a normal MD simulation, whereas a bias factor of $\gamma \rightarrow \infty$ ($\Delta T \rightarrow \infty$) recovers the standard Metadynamics.

The reconstruction of the free energy in multiple dimensions/along multiple reaction coordinates is also possible with Metadynamics. A further implementation, the Multiple-Walkers (MW)-Metadynamics¹⁵⁷ approach, allows the reconstruction of the free energy

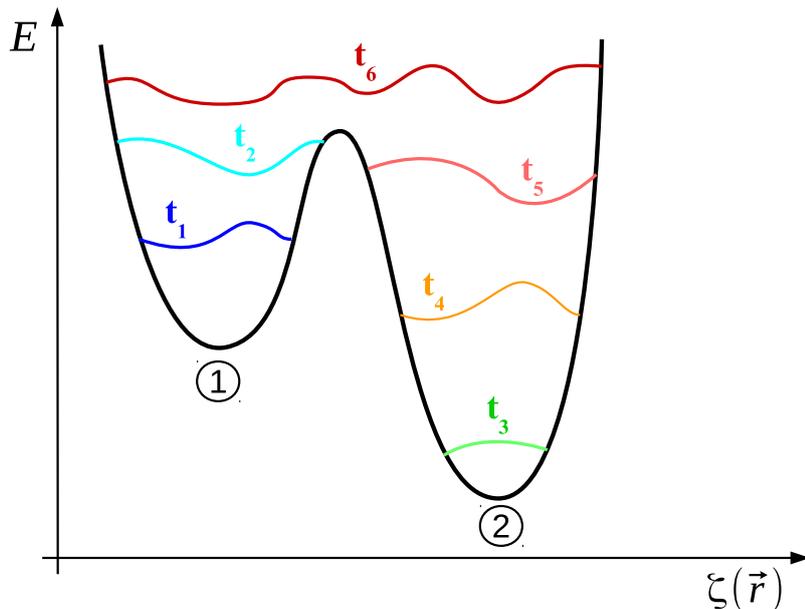


Figure 5.5: Enhanced sampling of the phase space with Metadynamics. A history-dependent bias potential $V(\zeta, t)$ acts on a predefined set of collective variables. The energy profile, which is a function of this collective variable set ζ , is continuously “filled up”. Consequently the system discourages “visited” phase space points and the sampling is enhanced in this specific collective variable space.

surface based on parallel simulations/walkers. In this work, the free energy along multiple reaction coordinates was reconstructed with MW-Metadynamics.

5.5 Proton Transfer Reaction Coordinates

Compared to other processes taking place in biomolecular systems, PT reactions are very fast and usually happen on a timescale of picoseconds. In such a case, it is possible to study the energetics of PT with unbiased multiscale MD simulations¹⁵⁸. However, for most PT reactions the barrier is significantly higher than $k_B T$ and the investigation of the transfer process requires enhanced sampling methods. Umbrella Sampling or Metadynamics are for this task a convenient choice and were also used in this work for the study of the long-range PT in the O state of bR. However, both of these methods require a predefined reaction coordinate which properly describes the essential degrees of freedom that rule the reaction kinetics. For a direct PT (e.g. between two amino acids or two water molecules), this is an easy task: One can use the anti-symmetric stretch coordinate $\xi = d^{D,H} - d^{A,H}$ which describes the charge travel simply on the basis of distances between the transferring proton with respect to its donor/acceptor. Unfortunately, such a collective variable is useless for the description of long-range PT reactions where the excess charge is transferred over multiple shuttling groups like in a water wire. The problem in this scenario is the rather concerted nature of the process, where the excess charge location is not well defined anymore, and multiple charge carriers (hydrogens) are available. Consequently, a proper descriptor for

long-range PT reactions may be based on the center of excess charge (CEC), and was already introduced by König et al.¹⁵⁹ as the modified center of excess charge (mCEC) coordinate:

$$\boldsymbol{\xi} = \sum_{i=1}^{N_H} \mathbf{r}^{H_i} - \sum_{j=1}^{N_X} w^{X_j} \mathbf{r}^{X_j} - \sum_{i=1}^{N_H} \sum_{j=1}^{N_X} f_{sw}(d^{H_i, X_j}) (\mathbf{r}^{H_i} - \mathbf{r}^{X_j}) \quad (5.11)$$

The first term in equation 5.11 is a sum over all hydrogen coordinates \mathbf{r}^{H_i} . The second term is a weighted sum over all positions \mathbf{r}^{X_j} of hydrogen coordinating atoms (e.g. oxygen, nitrogen etc.) taking part in the PT process, w^{X_j} is the number of protons/hydrogens coordinated to X_j in the least protonated configuration with respect to the reactant or product state. The last term can be seen as a correction which runs over all distances between hydrogens and coordinating atoms in order to decide, on basis of a switching function $f_{sw}(d^{H_i, X_j})$, which H_i and X_j atoms are connected by bonds:

$$f_{sw}(d^{H_i, X_j}) = \frac{1}{1 + \exp [(d^{H_i, X_j} - r_{sw}) / d_{sw}]} \quad (5.12)$$

r_{sw} and d_{sw} are empirical parameters which control the steepness and centering of the switching function on the bond length scale. The choice of these parameters will influence the transition state (TS) structures and the concertedness of the PT reaction.

With this, $\boldsymbol{\xi}$ is able to:

- reflect the center of excess charge (CEC) and the configuration/protonation state of all proton carriers.
- describe non-linear PT pathways.
- allow free energy simulations without assuming the PT pathway a priori.

However $\boldsymbol{\xi}$ defined like in equation 5.11 is a vector, but for Umbrella Sampling or Metadynamics we need a scalar quantity. This is accomplished by mapping $\boldsymbol{\xi}$ relative from an initial donor D to a final acceptor A :

$$\zeta_1 = \frac{d_{\boldsymbol{\xi}, D}}{d_{\boldsymbol{\xi}, D} + d_{\boldsymbol{\xi}, A}} \quad \text{or} \quad \zeta_2 = \frac{d_{\boldsymbol{\xi}, D} - d_{\boldsymbol{\xi}, A}}{d_{\boldsymbol{\xi}, D} + d_{\boldsymbol{\xi}, A}} \quad (5.13)$$

With any of these definitions, the CEC transfer is defined in a range from $\zeta_1 = 0/\zeta_2 = -1$, meaning the charge is located on the initial donor, to $\zeta_{1/2} = 1$, meaning the charge is located on the final acceptor.

Part III

Results

Structural Models for the bR/O/O* State

In order to resolve the energetics and the mechanism of the assumed long-range PT, taking place in the O→bR transition, it is indispensable to work out the structural properties of the individual states which are part of the PT process. Accordingly, in this chapter structural models for the bR, O and O* state will be proposed.

While the structure of the bR state is very well known and resolved with high-resolution, there is only few structural data for the O and none for the O* state available. Thus, one has to rely on structural models derived from MD simulations for these states. However, also in the theoretically derived models there are discrepancies in structural properties which are important for the PT process^a.

Consequently, in this chapter we aim to predict reliable structures for the two O states, by minimizing the sampling problem with the help of the enhanced sampling technique HREX. On basis of these simulations the most significant structural differences between the three models will be analyzed. Of special interest is the water distribution and the possible occurrence of a water wire in the region between D85 and the PRG, since this is a crucial prerequisite for the assumed long-range PT.

Furthermore, the focus will be on structural rearrangements of amino acids in this reactive region and their stable conformations with respect to Gibbs free energy.

6.1 Simulation Setup

As basis for the structural models of the bR, O and O* state served a recently released crystal structure of the bR state which was obtained by time-resolved serial femtosecond crystallography²⁵. The heavy atom coordinates (see Figure 6.1A) of this structure were obtained from the PDB file with the ID 5B6V. Beside the amino acid sequence, this file also contained 17 entries for crystal water which were not deleted. This is mainly because most of these water molecules were located in the active site cavity close to D85, D212,

^aFor example the observation of a water wire between D85 and the PRG^{68,69,160}.

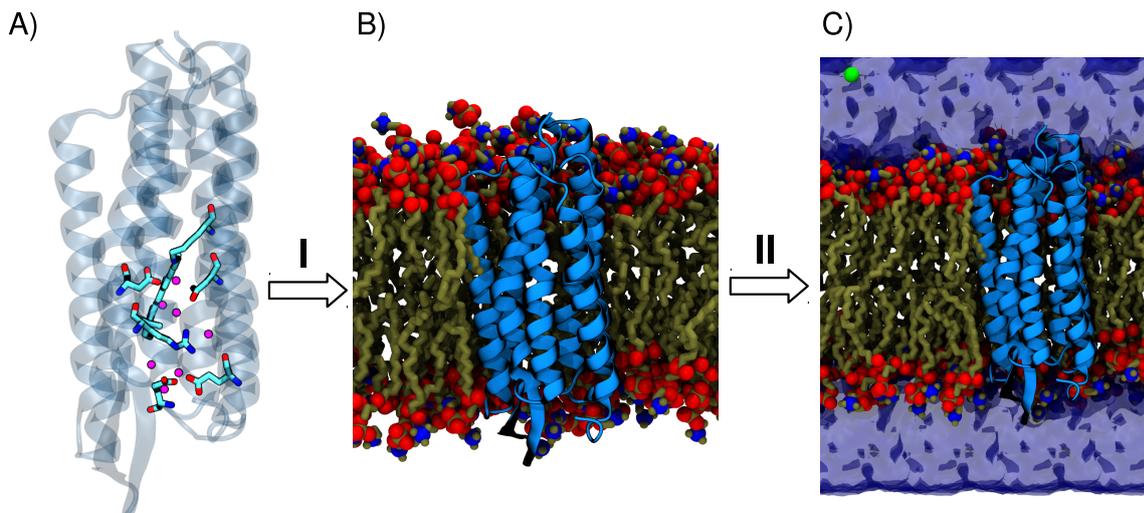


Figure 6.1: Workflow of the bR model setup. The crystal structure with the PDB-ID 5B6V (A) was embedded in a POPC bilayer (during step I). The protein/membrane-complex was subsequently solvated and a chloride ion was added in order to gain charge neutrality of the system (during step II). The whole system comprised the truncated 230 amino acid long, ground state bR, 284 POPC lipids, 16381 TIP3P waters and one chloride counterion (C).

R82, E204, E194 and it is well known that this region contains structurally important water molecules which build up a HBN. This decision is furthermore underpinned by the relatively high crystallographic resolution of 2.0 \AA and the very low, thus reliable, B-factors ($38.9\text{-}72.7 \text{ \AA}^2$) of the water oxygens.

For MD simulations, Gromacs v.5.0¹⁶¹ in combination with the Plumed v.2.1.1 plugin¹⁶² was used. The CHARMM36 force field was used for the description of the whole system comprising protein, lipid phase, water phase and ions. Parameters for the retinal molecule were taken from refs. 163–167. During topology creation with *pdb2gmx* all titratable amino acids were kept in their physiological protonation state, identical to the bR state. As lipid environment, a 500 ns preequilibrated POPC bilayer was taken which was provided by Violetta Schneider.

For the embedding of the bR protein into the POPC bilayer the InflateGRO¹⁶⁸ methodology was used. During this workflow the protein was centered in the lipid bilayer. In the next step, overlapping lipids within a cutoff radius of 14 \AA were removed, and the bilayer was artificially inflated in its plane. At the end, several shrinking steps followed which compressed the bilayer and packed the lipids around the protein. The shrinking procedure was seen as “converged”, when the area per lipid value of the POPC bilayer was close to the experimental value¹⁶⁹ of 65.8 \AA^2 (see Figure A.1).

The protein/membrane-complex (see Figure 6.1B) was solvated by using the *gmx solvate* routine with the CHARMM TIP3P as water model. During solvation, small gaps between the lipid acyl chains also got filled with water. These misplaced water molecules were removed from the structure with the help of a Python script which was written by Dr.

Sabine ReiSSer. In order to gain charge neutrality of the system one chloride counterion was added with the *gmx genion* routine. The whole workflow of the system setup is summarized in Figure 6.1. The whole system comprised the truncated 230 amino acid long, ground state bR, 284 POPC lipids, 16381 TIP3P waters and one chloride counterion (see Figure 6.1C).

Energy minimization was conducted using the steepest descent minimization algorithm with a maximum force threshold of 1000 kJ/(mol · nm). The system was equilibrated in the NVT ensemble for 200 ps at 300 K using the V-rescale¹⁷⁰ thermostat. Furthermore, xyz-position restraints were placed on the heavy atoms of protein and lipids. An unrestrained NPT simulation over 100 ns followed with a temperature and pressure coupling to 300 K and 1.0 bar. The temperature was controlled via the Nosé-Hoover thermostat and the pressure via the Parrinello-Rahman barostat. Since it is known that MD simulations are insufficient for the preservation of the pentagonal HBN⁶⁷, an additional bias potential was applied which should ensure the intactness of the HBN. Therefore, distance-dependent, harmonic restraints were applied between the HBN water molecules and D85, D212, K216RET. The biasing force of $\kappa = 8000 \text{ kJ}/(\text{mol}\cdot\text{nm}^2)$ was activated, when the distance between the water oxygen atom and the respective amino acid atom was larger than 2.8 Å.

6.1.1 Protonation States and Hamiltonian Replica Exchange Simulations

On basis of the 100 ns equilibrated bR state structure the two further models for the O and O* state were created. Therefore, protonation states of the aspartates/glutamates were changed. More precisely, for the O state model, E204 was deprotonated and D85 protonated. The same was done to mimic the O* state but instead of D85, D212 was protonated. The O and O* state model were minimized and NVT equilibrated like before (cf. section 6.1, p 47).

The three protonation models were then prepared for enhanced sampling simulations employing HREX¹⁵¹. Since it is known that the NPT ensemble disrupts the HREX exchange rates, these simulations were conducted in the NVT ensemble with the V-rescale thermostat. All three protonation models were equilibrated over 200 ps. The HREX simulations had for each protonation model the same setup: 16 replicas were used which spanned a λ scaling range from 1 – 0.16. This corresponds to a simulation temperature range of 300 – 1900 K. The frequency for an exchange-attempt was set to 4 ps. The region, where the scaling was applied to, consisted of the active site amino acids D85, D212, K216RET, R82, E194, E204 and all interacting amino acids in and around this region. A complete list with the amino acids of the “hot region” and the justification for their selection is shown in Table 6.1. The finished simulations were then analyzed for a proper exchange of the individual replicas. All three simulations exhibited a frequent exchange of their replicas (see Figure A.2) which is a prerequisite for efficient sampling. As basis for the analysis of each state, served the corresponding unscaled ($\lambda = 1$) replica.

Table 6.1: Selected amino acids of the HREX “hot region” are listed together with their justification of selection.

Amino acid	Function
D85	proton donor + member of the pentagonal HBN
D212	proton acceptor/donor + member of the pentagonal HBN
K216RET	Member of the pentagonal HBN
R82	Up-/Downswing movement during photocycle
E194	Part of the PRG region
E204	Part of the PRG region
Y185	Hydrogen-bonded to D212
W86	Hydrogen-bonded to D212
E9	Polar amino acid near to the PRG
Y83	Polar amino acid near to the PRG
S193	Hydrogen-bonded to the E204
P77	Polar amino acid near to the PRG + hydrogen-bonded to S193
Y57	Hydrogen-bonded to D212
F208	Serves together with R82 as hydrophobic plug
Y79	Polar amino acid near to the PRG + hydrogen-bonded to E9
P200	Hydrogen-bonded to E204
W189	Polar amino acid near to the PRG + hydrogen-bonded to Y83

6.2 Key Differences between the bR/O/O* State

6.2.1 Dynamics of the R82 Side Chain

During the 10 ns of HREX simulation, the three bR models diverge from each other with respect to the active site geometry. Probably the most significant structural rearrangements between the bR state and the two O states undergoes the arginine side chain R82. First, its reorientation towards the extracellular or cytoplasmic side will be addressed.

This movement of R82 was already observed by Clemens et al. in an “early M-like” state of bR⁴⁴. In order to describe this “swing” movement, we use the distance between the stiff nitrogen atom of A44 and the C ζ atom of R82 (see Figure 6.2). This collective variable is very similar to the one that Clemens et al. used. Figure 6.3A displays the R82-A44 distance for all three protonation models over 10 ns of HREX simulation. For the bR state simulation, the R82-A44 distance stays over the whole simulation at around 2.3 nm. This means, the R82 side chain is oriented towards the cytoplasmic side. Hence, R82 is close to D85/D212 (also called “upper” active site) and interacts with D212 (like in Figure 6.2). For the two O state models this is different: The R82-A44 distance has already increased during NVT equilibration (O* state) or within 500 ps of HREX (O state) to around 2.8 nm. The distance stays for the whole simulation time at this value. For the O and O* structure this means that the R82 side chain is oriented towards the extracellular side with E194/E204. The arginine is hence close to the “lower” active site and interacts with the PRG (see O state models in Figure 6.9). The histograms in Figure 6.3B were created with the

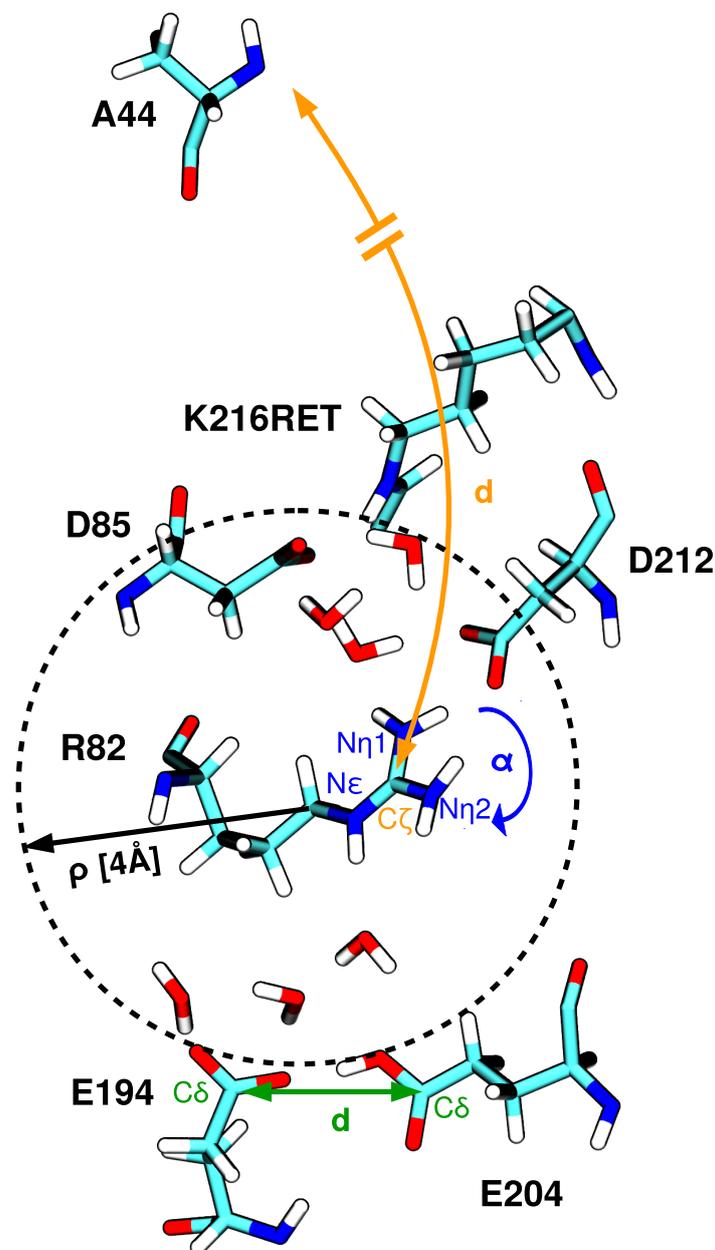


Figure 6.2: Summary of the collective variables that were used for analysis of HREX simulations or in Umbrella Sampling. The PRG gate movement is represented by the distance between the two C δ atoms of E194 and E204 (green). As measure for the cavity hydration the number of water molecules within 4Å of the R82 residue is used (black). The R82 swing movement is described by the distance between the R82 C ζ atom and the backbone nitrogen of the stiff A44 residue (orange). In order to monitor flip movements of the R82 guanidinium group an internal dihedral angle α was defined (blue).

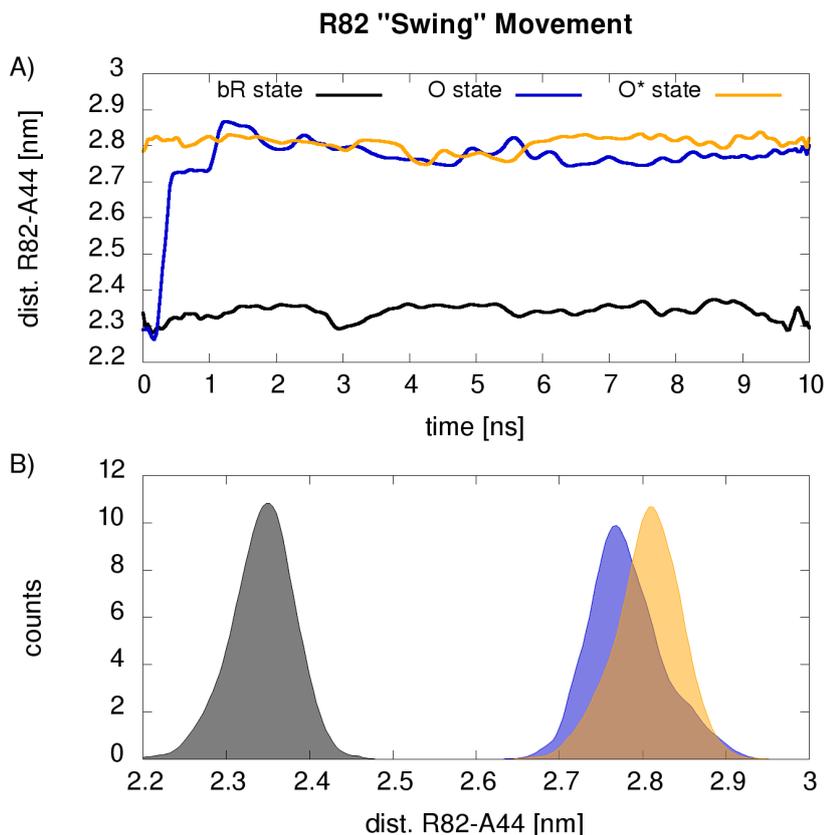


Figure 6.3: A: The distance between the A44N and the R82C ζ atom is displayed over the whole 10 ns of HREX sampling for the bR (black line), O (blue line) and O* (yellow line) state. B: Histograms for the A44N-R82C ζ distance. The histogram is based on the last nine nanoseconds of bR (black area), O (blue area) and O* (yellow area) state HREX simulation.

data over the last nine nanoseconds and clarify that each of the three protonation models exhibits only one preferred R82 orientation. Moreover, the histograms illustrate again that the two O states are similar concerning their R82 orientation. The second structural rearrangement that happens during the O \rightarrow bR transition is an internal "flip" movement of the R82 guanidinium group. In the bR state, the N ϵ and N η 1 atoms of R82 are oriented towards the PRG, whereas the N η 2 atom is facing D212. In all three HREX simulations, flip movements by ca. 180° of the planar guanidinium group can be observed such that the orientation of the mentioned atoms are interchanged. For the purpose of describing this movement, an internal dihedral angle named α , defined by the R82 atoms H α -C α -N ϵ -H ϵ , was used (see Figure 6.2). Figure 6.4A shows the orientation of the guanidinium group over the whole 10 ns of HREX sampling for all three bR models. In the bR state simulation the guanidinium group stays constantly in the $\alpha = 2.5$ rad/145° orientation. After around 7 ns a few flip movements can be observed into the conformation with $\alpha = -1.13$ rad/-65°. The analysis of the O* state trajectory shows a similar stable guanidinium conformation:

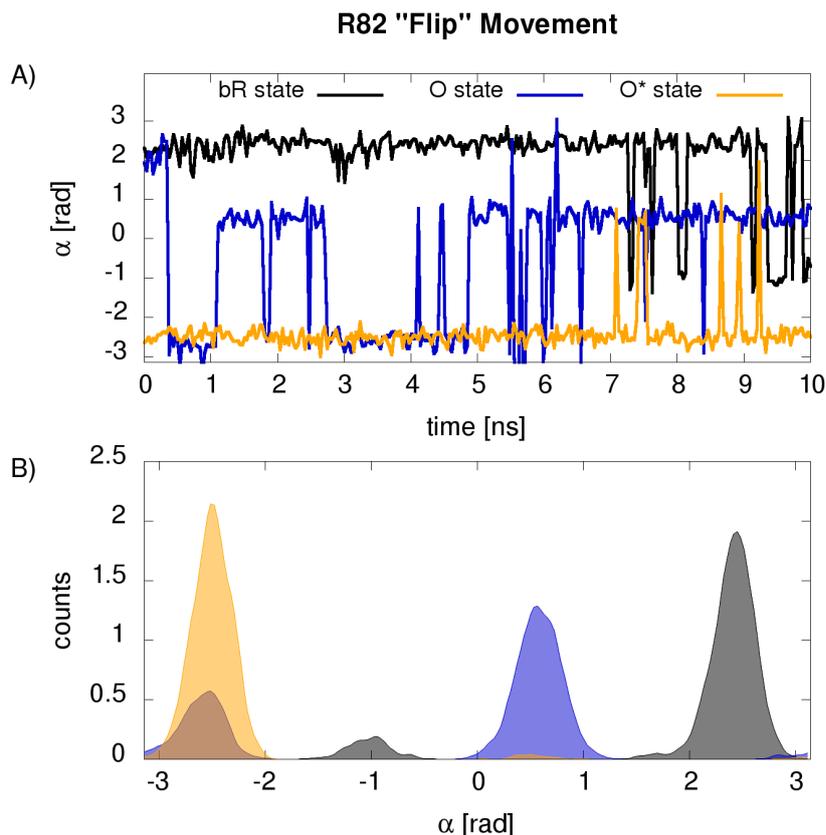


Figure 6.4: A: The orientation of the planar R82 guanidinium group is displayed (via α) over the whole 10 ns of HREX sampling for the bR (black line), O (blue line) and O* (yellow line) state. B: Histograms for the guanidinium group orientation. The data is based on the last nine nanoseconds of the bR (black area), O (blue area) and O* (yellow area) state HREX simulation.

For the first 7 ns the planar group stays in its $\alpha = -2.5$ rad/ -145° conformation and then begins shortly to populate the $\alpha = -0.5$ rad/ 28° orientation. Contrary to the bR and O* state, the guanidinium orientation fluctuates much more in the O state simulation. After an equilibration time of around 1 ns, the guanidinium group populates its most favorable conformation with $\alpha = 0.5$ rad/ 28° . However, the guanidinium group flips frequently to the conformation with $\alpha = -2.5$ rad/ -145° . Figure 6.4B shows for each of the three HREX simulations the corresponding histogram for the α angle. The histograms are based on the last nine nanoseconds of simulation. Again, it can be observed that the R82 guanidinium group populates in the bR and O* state mainly one conformation, whereas in the O state the planar group also significantly populates a second conformation. Based on these results, the most stable conformation of the guanidinium group is for the bR and O* state the one where the $N\epsilon$ and $N\eta_1$ atoms are oriented towards the PRG. This is different for the O state. The most favorable guanidinium orientation for this model is the one, where the $N\eta_2$ atom is facing the PRG (see Figure 6.9).

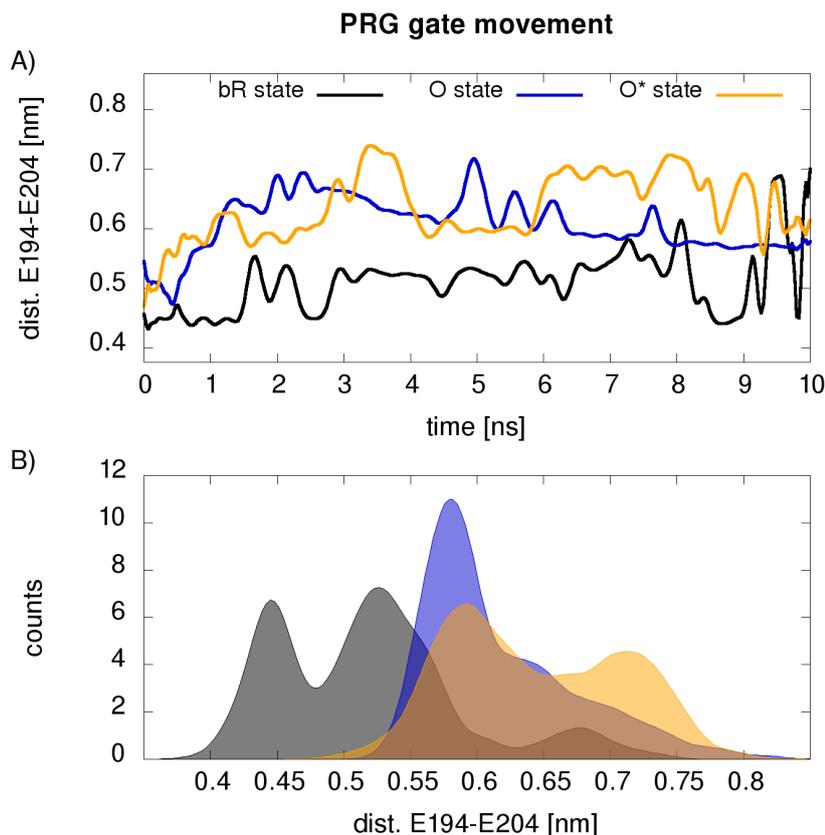


Figure 6.5: A: The dynamics of the PRG is displayed via the E204C δ -E194C δ distance over 10 ns of HREX sampling for the bR (black line), O (blue line) and O* (yellow line) state. B: Histograms for the E204C δ -E194C δ distance. The data is based on the last nine nanoseconds of the bR (black area), O (blue area) and O* (yellow area) state HREX simulation.

6.2.2 Dynamics of the PRG

The modification of the protonation states in the three different models also affects the conformation of the PRG amino acids E194 and E204. In the bR state, the O ϵ 2 atom of E204 is protonated and hence able to form a hydrogen bond to one of the E194 oxygens. This ability gets lost in the O or O* state where E204 is deprotonated. In order to display the conformational changes introduced by the deprotonation, the distance between the E204-C δ and E194-C δ atom serves as an indicator (see Figure 6.2). Figure 6.5A displays for all three protonation models this distance over the whole range of 10 ns HREX sampling. The corresponding histogram plot (see Figure 6.5B) shows that the bR state structure stays nearly the whole simulation time in two conformations where the E194C δ -E204C δ distance is around 0.45 and 0.55 nm. Both structures are populated nearly equally with a slight preference for the conformation with the larger glutamate distance. In the underlying conformation, E194 is turned away from E204 and hydrogen bonds to Y83, whereas E204 is stabilized by a hydrogen bond to the backbone NH group of E194. The second prominent

geometry, where E194C δ -E204C δ distance is decreased, also exhibits these interactions but furthermore E204 and E194 are hydrogen-bonded to each other. For the two O state models, the glutamate distance increases directly over the first 1.5 ns. This seems reasonable since there is no hydrogen bond anymore between the two glutamates. In the O state simulation, one main conformation is populated where the E194C δ -E204C δ distance is ca. 0.6 nm. This conformation also gets primarily populated in the O* state simulation. In the underlying structure, E194 is stabilized by a hydrogen bond to Y83 and the other glutamate E204 is interacting with R82 over a salt bridge. Furthermore, the O* state populates less frequently a second conformation with 0.72 nm. In this structure, the E194C δ -E204C δ distance is even more increased so that the E194 is hydrogen-bonded to Y83 and salt-bridged to R82.

6.2.3 Hydration of the Active Site Cavity

Beside the mentioned structural rearrangements of amino acids, there also occurs during HREX sampling in some protonation models a change in the internal protein hydration. The reason therefore is the presence of a cavity near to the extracellular “exit” of bR which can change its hydration level. This cavity is located between the upper and lower active site. Directly in the middle of this cavity lies the R82 side chain. Hence, as indicator for the cavity hydration, the number of water molecules within 4Å of the R82 was taken (see Figure 6.2). Figure 6.6A depicts the cavity hydration over the 10 ns of HREX sampling for the bR, O and O* state. In the bR state, already seven water molecules were present at the beginning of the simulation. This is due to the 100 ns of NPT equilibration and the presence of crystal water molecules which were not deleted during model setup. In the course of the HREX simulation, the cavity hydration stays nearly the same. In the bR state, mainly 6 – 7 water molecules are favored in this cavity (see Figure 6.6B). For the two O state models, the cavity hydration is at the beginning gradually increased and at the end nearly twice as large as for the bR state. The corresponding histograms show that for the O* state model mainly ten water molecules are present in the cavity. For the O state model the preferred hydration is a little bit larger with eleven water molecules. Figure 6.6A also shows that the filling up of this cavity happens on a time scale of nanoseconds, which is in the context of water dynamics a very slow process. Because of that, the histograms in Figure 6.6 are based on the last seven nanoseconds of the HREX data.

A comparison with the results for the PRG dynamics reveals that the filling up of the cavity is nearly on the same timescale as the increase of the E204C δ -E194C δ distance. Hence, it is probable that these processes are correlated and the increase of the PRG distance is like a gate opening that ensures the influx of extracellular water molecules. Moreover, since the R82 side chain is located in the center of this cavity, it is likely that it also influences the cavity hydration. Let us have a closer look on the underlying structures in order to give a reason why the cavity hydration differs in the individual models. In the bR state, the PRG gate is mainly in its “closed” conformation and the R82 side chain is swung up towards the D85/D212 region. As a consequence, the hydration level stays constant and the cavity gets spatially reduced so that only water molecules are present in the lower region around the PRG group. This is different for the two O state models, where the PRG

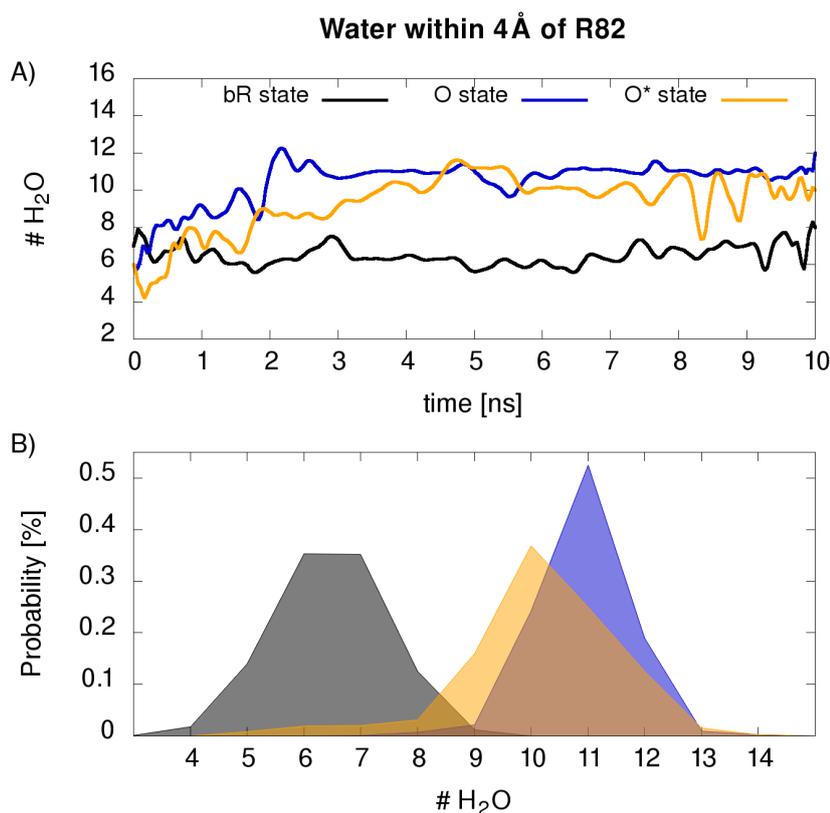


Figure 6.6: A: The internal cavity hydration of the bR (black line), O (blue line) and O* (yellow line) state is shown over 10 ns of HREX sampling. As indicator for the cavity hydration all water molecules within 4 Å of R82 are counted. B: Histograms for the cavity hydration based on the last seven nanoseconds of the HREX simulations. The histogram for the bR, O and O* state is depicted as black, blue and yellow area respectively.

gate is in its “open” conformation and the R82 side chain is swung down towards the PRG. Mainly these two properties cause the increase of the cavity hydration: Extracellular water molecules can enter and furthermore are able to spread out over the whole cavity region. Hence, only in the simulations for the two O state models water wires that connect the upper D85 with the lower PRG site can be observed. Furthermore, the guanidinium group plays with its five hydrogen bond donation sites an essential role. This property qualifies the side chain to organize an ordered water network between the upper and lower active site. Moreover, this ordered HBN could be altered through the “flip” movement of this functional group.

6.3 PMF Calculations for the R82 Side Chain

The previous results have shown that the R82 side chain is probably the most flexible amino acid in the cavity region, and furthermore dictates primarily its hydration in combination with the PRG. This task is ensured by the two principal dynamics of the R82 side chain: the swing and flip movement.

In this section the most stable conformations with respect to these two movements should be identified for each of the three bR protonation states. Furthermore, with this analysis, other possible R82 conformations should be found and their difference in free energy evaluated. In order to achieve this, Umbrella Sampling was performed on the swing and flip collective variable for each protonation state.

6.3.1 Simulation Setup

The A44N-R82C ζ distance served as reaction coordinate for the swing movement. This distance was scanned from 2.1 – 2.9 nm over 15 windows (0.06 nm displacement). In each window, a harmonic potential was applied with a spring constant of $\kappa = 4000$ kJ/(mol·nm²). Each window was simulated in the NPT ensemble at 300 K (Nosé-Hoover) and 1.0 bar (Parinello-Rahman) for 10 ns. The starting structure for each state was based on the corresponding HREX end structure. The last eight nanoseconds of Umbrella Sampling were used for statistical evaluation. In order to check for a sufficient overlap between the individual windows, histograms were created with *g_analyze*. The calculation of the PMF was done with WHAM¹⁵³. Therefore, the implementation of Alan Grossfield¹⁷¹ was used.

In order to scan the free energy profile of the R82 flip movement an internal dihedral angle α was used, defined by the R82 atoms H α -C α -C ζ -N η 1^b. The dihedral angle α was scanned from $-\pi$ to π over 42 windows (0.15 rad displacement). In every window, a harmonic spring constant of $\kappa = 800$ kJ/(mol·rad²) was applied. The perturbation of the planarity of the guanidinium group is a delicate issue for force fields. The atoms N ϵ , C ζ , N η 1 and N η 2 are lying in one plane and the positive charge is delocalized between them. Force fields aren't able to describe quantum effects like delocalization. Since the α angle directly perturbs the geometry of the guanidinium plane, a QM description of this residue is indispensable. Hence, the Umbrella Sampling simulations for the "flip" movement were performed with QM/MM. The QM region comprised the R82 side chain from C β on and was treated with DFTB3/3OB with D3BJ dispersion correction. The MM part was described with the CHARMM36 force field. Each window was sampled for 1.7 ns at 300 K (Nosé-Hoover) and 1.0 bar (Parinello-Rahman). The last 1.2 ns of Umbrella Sampling were used for the calculation of the PMF with WHAM.

The histograms for the swing reaction coordinate show for each bR model sufficient overlap (see Figure A.3). The resulting PMF plots for this structural rearrangement are shown in Figure 6.7. For the bR state, two minimum conformations are identifiable. The global minimum forms the conformation where the R82 side chain is in the swung-up

^bThis reaction coordinate is slightly different to the one used for the analysis of the HREX simulations in subsection 6.2.1.

position with a A44N-R82C ζ distance of around 2.35 nm. The second minimum forms the conformation where this amino acid is in the swung-down position with a A44N-R82C ζ distance of ca. 2.7 nm. On the free energy profile, the latter conformation lies 8.4 kcal/mol higher than the global minimum. Furthermore, the swung-down conformation seems to be a metastable state since the barrier height for this process lies only 1.3 kcal/mol higher. The preference for the swung-up R82 conformation in the bR state is due to the favorable hydrogen-bonding situation in this state: The guanidinium group forms together with the side chains of D212 and Y57 a complex which is strongly stabilized by three hydrogen bonds. Contrary to this, the swung-down conformation is only loosely stabilized by fluctuating hydrogen bonds of the R82 side chain with the E194 and T205 side chains.

For the two O state models there exists only one minimum: the R82 in the swung-down conformation. In the O state PMF this global minimum is localized at a A44N-R82C ζ distance of around 2.75 nm, whereas it is shifted a little to 2.8 nm in the PMF of the O* state. Again, the hydrogen bonding situation is responsible for these results. In the O state, the D85 is protonated and forms with the D212 and surrounding water molecules a HBN. Consequently, the R82 guanidinium group has in the upper active region no hydrogen bond acceptor anymore. Contrary to this, in the swung-down conformation the R82 side chain is stabilized by hydrogen bonds to the E204 side chain and the E194 side chain is primarily hydrogen-bonded to the OH group of Y83. Nearly the same situation applies for the O* state: The D212 is protonated and hence can not stabilize the R82 side chain, yielding to no minimum on the free energy profile for the swung-up conformation. Instead, the guanidinium group is even a little bit further oriented towards the PRG as in the O state so that it could form hydrogen bonds with E204 and E194. These results are in accordance with the populated conformations of the HREX sampling (see Figure 6.3).

Also for the flip movement, the Umbrella sampling windows exhibited a reasonable overlap (see Figure A.4). The resulting PMF of this movement shows for the bR state two minima. The global minimum is located at $\alpha = -0.85$ rad. For further discussion this conformation is called the non-flipped conformation. In the non-flipped bR state conformation the N η 1 atom is facing the upper active site and the N ϵ atom is facing the PRG. This conformation forms the global minimum on the free energy profile because the guanidinium plane is perfectly oriented such that the guanidinium N η 1 and NH η 2 atom can form the hydrogen-bonded complex with the D212 and T57 side chain. Furthermore, the other unoccupied hydrogen bond acceptors can interact with the water molecules in the lower cavity region. For the bR state there exists another minimum, where the guanidinium group is flipped by nearly 180° ($\alpha = \pi$ rad). The flipped bR state minimum lies ca. 2.9 kcal/mol higher on the free energy profile compared to the global minimum. In this conformation the R82 guanidinium group is not stabilized by hydrogen bonds to other amino acid side chains. Instead, the guanidinium plane is solvated and forms hydrogen bonds to the cavity water molecules. The minimum free energy path that connects the non-flipped and flipped bR state basin amounts to ca. 6.2 kcal/mol.

In the O state PMF the global minimum is found at an α angle of $-\pi$ rad. Thus, contrary to the bR state, in the O state the flipped conformation forms the most stable one. The second minimum on the free energy profile can be found at $\alpha = 1.0$, where the guanidinium

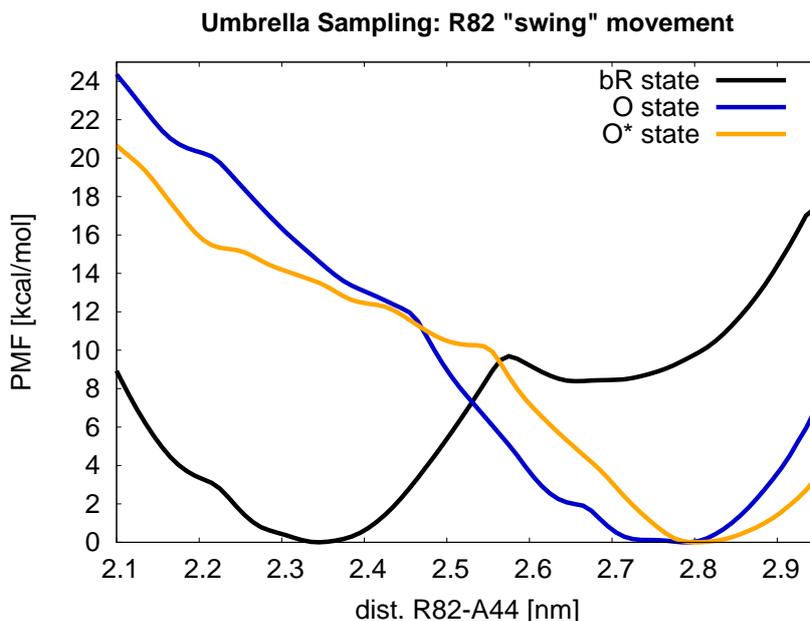


Figure 6.7: Potential of Mean Force for the R82 “swing” movement. As reaction coordinate the A44N-R82C ζ distance was used. The minimum at 2.35 nm represents the R82 in the swung-up conformation, whereas the minima at 2.7–2.8 nm represents the side chain in its swung-down state.

group is in its non-flipped conformation. This minimum lies only 0.85 kcal/mol higher than the flipped conformation. Compared to the bR state, the α angle for this conformation is shifted by ca. $106^\circ/1.85$ rad. The minimal reaction barrier between the flipped and non-flipped O conformation amounts to ca. 6.0 kcal/mol. A look on the underlying structures of both minimum conformations, reveals that in the flipped O conformation the N η 1 atom of the guanidinium group is strongly bound via a salt bridge to the PRG amino acids E194/E204. Furthermore, the N ϵ atom serves as hydrogen bond donor for a rigid water molecule close to Y83. Moreover, the NH2 atom interacts with the cavity water molecules. In the non-flipped O conformation these interactions are weakened: The N η 2 atom forms a salt bridge to E204 and the two other guanidinium nitrogens, that are facing D212, interact with the cavity water molecules.

For the O* state the most stable guanidinium flip orientation is, like in the bR state, the towards 0.6 rad shifted non-flipped one. This conformation is with respect to Figure 6.8 1.1 kcal/mol lower in energy than the second lowest minimum at $\alpha = -2.5$ rad which represents the flipped O* conformation. The minimum free energy path that connects both valleys amounts to 7.0 kcal/mol. Like for the O state, the free energy difference ΔG between the flipped and non-flipped conformation is very small. In the two corresponding O* state conformations, the nitrogen atoms of the guanidinium group interact with the E204 side chain, the rigid water molecule close to Y83 and with the water molecules in the upper cavity region close to D212.

The identified minimum conformations of the three protonation states were already

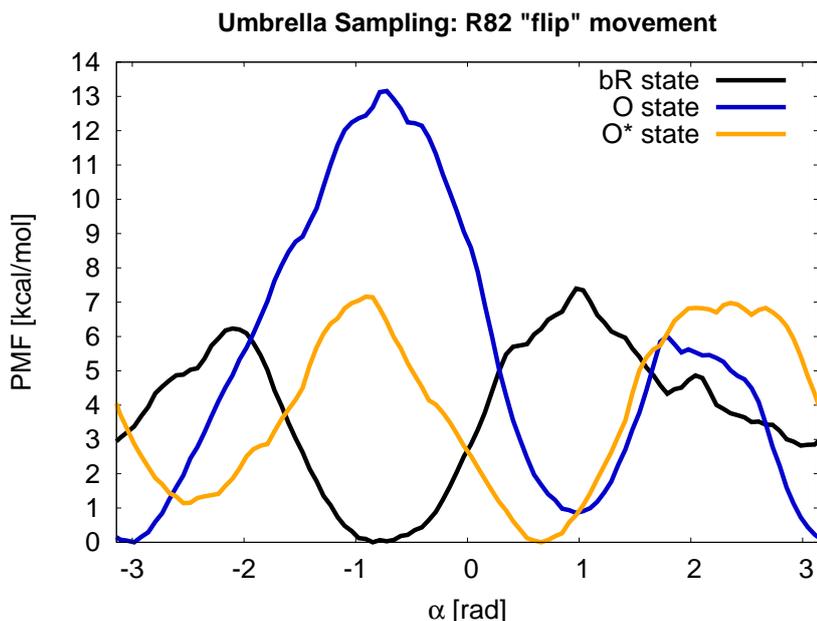


Figure 6.8: Potential of Mean Force for the R82 flip movement. The internal R82 dihedral angle α defined by the atoms $\text{H}\alpha\text{-C}\alpha\text{-C}\zeta\text{-N}\eta 1$ was used as reaction coordinate. The minimum conformations around $\alpha = \pm\pi$ rad represent the R82 guanidinium group in its “flipped” conformation. In the other minima at $\alpha = \pm 0.85$ and 1.0 rad the guanidinium group is in the “non-flipped” conformation.

observed in the corresponding HREX simulations in section 6.2.1. Keep in mind that the definition of α in this section is directly inverted compared to the one used for the HREX analysis.

6.4 Conclusions

The first part of this chapter outlined how the bR, O, and O* state model of bacteriorhodopsin were derived, based on a time-resolved crystal structure of the bR state. Beside conventional force field methods, also HREX was employed in order to ensure for sufficient sampling of the O state active site region. Furthermore, the main structural rearrangements in this region have been identified and analyzed for all three models.

The most significant changes between the protonation models undergo the R82 side chain and the PRG consisting of E194 and E204. These charged amino acids react very sensible because a change in the protonation pattern involves for these amino acids an extensive change in the hydrogen-bonding/salt-bridging situation. Moreover, it has been shown that the R82 and PRG dynamics also influence the hydration level of the internal cavity between D85/D212 and the PRG. Upon protonation change, the PRG reacts like a gate. In the bR state, the E204 is protonated and the gate is closed due to a hydrogen bond between E204 and E194 and no extracellular water molecules can enter the cavity. The gate gets opened, and hence water can flow in, after deprotonation of E204 – in the O

or O* state – due to the lack of the hydrogen bond between E194 and E204. The R82 side chain reacts upon a protonation change like a plug that blocks or opens the entrance for water molecules to the upper active site close to D85/D212. When the PRG is protonated, like in the bR state, the “plug” is in its swung-up conformation (in interaction with D212) and blocks the entrance to the upper active site. Since the PRG is also closed in this state, only preexisting cavity water molecules are present in the lower cavity region. Hence, the cavity hydration level for the bR state is low. Contrary to this, in the two O states the PRG gate is deprotonated, open and forms salt bridges with the R82 side chain. Thus, the plug is in its swung-down conformation. As a consequence, extracellular water molecules can enter the cavity and spread out over the whole cavity also into the upper active site region. Hence, the cavity hydration level for the O states is high.

Another structural rearrangement that was observed between the three protonation states, was the flip movement of the R82 guanidinium group. With respect to the PMF calculations, the bR and O* state favor the non-flipped conformation, whereas the O state favors the flipped conformation. However, it has to be mentioned that the energetical differences between these conformations are very small (1–3 kcal/mol) and the flip movement plays, compared to the dynamics of the PRG gate and R82 swing movement, a minor role. Nevertheless, the guanidinium group is a crucial component for the existence of the observed water wires in the two O state models. In these states, the NH groups of the swung-down guanidinium group serve as hydrogen bond donors for the cavity water molecules and hence arrange a water wire between D85 and the PRG. Thus, the R82 probably prepares the water wire for the long-range PT from D85 to E204^c.

The results have shown that the O and O* model are very similar to each other concerning cavity hydration level and R82/PRG conformation, whereas the models differ exactly in these properties from the bR state.

^cThe mentioned properties of the arginine side chain were also observed by Armstrong et al. in voltage sensitive ion channels¹⁷². There the plug movement of arginine controls the passage of ions through the channel and its movement is also controlled by hydrogen bonding of the guanidinium plane with polar amino acids.

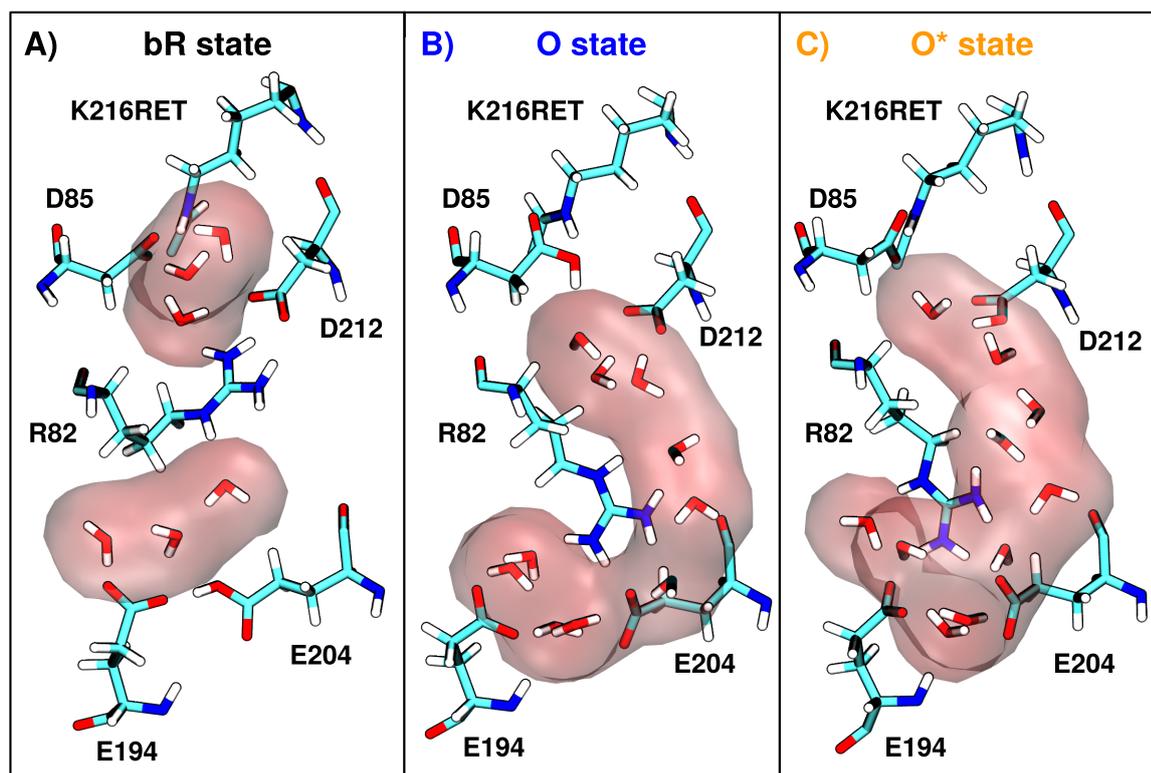


Figure 6.9: HREX sampled structures of the active site for the bR (A), O (B) and O* state (C). In the bR state the R82 side chain is swung up and non-flipped. The PRG gate is closed. Hence, the hydration level of the cavity is low and the connection to the D85/D212 site is interrupted by R82. This is different for the O and O* state. The R82 is in its swung-down conformation and the PRG gate is open. Moreover, in the O state the R82 guanidinium group is flipped. As a consequence, the water density in the cavity is raised and the D85/D212 region is connected via a water wire with the PRG region.

O→O*: A Direct Proton Transfer

With respect to the resolved structural properties of the O and O* state, one can assume that the O→bR transition probably involves only little changes in the bR active site.

However, a MM-based description of the O* state tells nothing about the stability of the transferring proton in this state. In order to address this question, in this chapter the description of the upper active site is switched to a DFTB3-based QM/MM setup. The subsequent prediction of Gibbs free energies for the PT in the O→O* transition should provide a quantitative measure of the PT product stability.

The reaction barrier for the O→O* transition is of further special interest since it is assumed that the deprotonation of the D85 represents the rate-limiting step in the O→bR transition. Hence, a prediction of $\Delta G_{O-O^*}^\ddagger$ would allow the comparison with experimental estimates for the O→bR transition.

Beside the energetics of the PT reaction, the focus will be placed on possible PT pathways and their mechanisms.

7.1 Stability of the Reactant and Product State

7.1.1 Simulation Setup

A structure from the O state HREX simulation around 9.95 ns served as starting structure for the simulation of the reactant state. This geometry owns the in Chapter 6 mentioned key properties of the O state and is comparable to the in Figure 6.9B displayed structure. Since the PT product (O* state) is very similar to the O state, we took for the product starting structure the same as for the reactant and only changed protonation states (deprotonated D85, protonated D212).

The QM/MM simulations were conducted with a Gromacs-DFTB3 version¹⁷³ in combination with the Plumed v.2.1.1 plugin. For reactant and product, the QM/MM simulations and the equilibration were analogously set up: The QM region comprised the complete amino acid side chains of D85, D212 and K216RET from the C β atoms on. Furthermore, the starting structures contained four water molecules in the proximity of the three mentioned amino acids which were also included in the QM region. In total, the QM zone

comprised 90 atoms that were treated with DFTB3 in combination with the 3OBw⁹⁵ parameters. As dispersion correction, the BJ implementation of D3 was used. The rest of the system was described with the CHARMM36 forcefield parameters. An NPT equilibration followed for 2 ps at 300 K and 1.0 bar. For the temperature and pressure coupling, the Nosé-Hoover thermostat and the Parrinello-Rahman barostat were used. A production simulation over 200 ps in the NPT ensemble at 1.0 bar followed. In this simulation, Simulated Annealing was applied in order to increase the sampling of the two reactant states. The system was heated from 300 – 340 K over the first 30 ps. The temperature stayed at 340 K for the next 40 ps and was then decreased back to 300 K within 30 ps. The last 100 ps were conducted at 300 K. Moreover, to prevent the leakage of the QM-treated water molecules out of the QM region^a, spherical, harmonic position restraints were applied to the oxygen atoms of the four water molecules. Thereby, the oxygen atoms were allowed to travel freely from their initial positions in a sphere with 3.0 Å radius. If the distance to the spherical center was larger than 3.0 Å, a harmonical restraint with a spring constant of $\kappa = 2000$ kJ/(mol·nm²) set in and pulled the oxygens back towards their center. For evaluation of the two simulations, the last 100 ps of the corresponding trajectories were used.

7.1.2 Destruction and Formation of the Hydrogen-Bonded Network

The starting structure for the reactant state still exhibits after HREX sampling the pentagonal HBN that is formed by D85, D212, K216RET and three water molecules^b. Figure 7.1, left shows a geometry that is sampled frequently during the 100 ps of QM/MM. This conformation still exhibits the structural main features of the pentagonal HBN and is similar to the O state HBN structure, proposed by Phatak et al.⁶⁷. The only small deviation is the absence of the hydrogen bond between the to K216RET coordinated water molecule with D85. In the presented structure the water molecule only forms hydrogen bonds with the carboxy group of D212. Figure 7.1, right shows the same active site but with a representative, QM/MM sampled geometry of the PT product state. It is noticeable that during sampling, the pentagonal HBN broke down. Compared to the O state, the three hydrogen bonds (formed by two water molecules) which connected the carboxy groups of D85 and D212 with each other vanished. Instead, there is only one water molecule left in the product state that connects the two aspartates over two hydrogen bonds. Moreover, the Schiff base NH group lost its hydrogen bond to the water molecule and instead is tilted and hydrogen-bonded to one oxygen atom of the D85 carboxy group.

The pentagonal HBN network serves like a scaffold which forms a gap between the two aspartates and prevents the direct interaction of the two amino acids. Thus, the break down of the HBN leads to further small structural rearrangements in the upper active site. The most significant impact is the spatial approach of the two aspartates D85 and D212. This is validated by Figure 7.2 which shows histograms of the D85C γ -D212C γ distance, evaluated for the last 100 ps of QM/MM. The Figure displays for the aspartate distance a separated

^aOr the influx of MM water molecules into the QM region.

^bThe preservation of the HBN is probably due to the applied position restraints during the HREX equilibration (cf. Section 6.1, p.47).

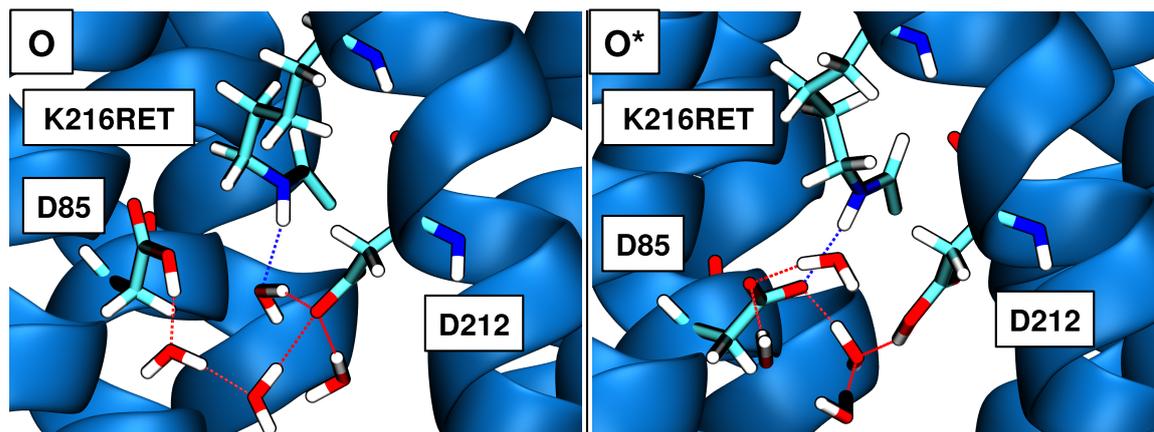


Figure 7.1: O (left) and O* (right) state geometries of the upper active site after 200 ps of QM/MM simulation in combination with Simulated Annealing. In the O state the pentagonal HBN is preserved, whereas it is disrupted in the O* state.

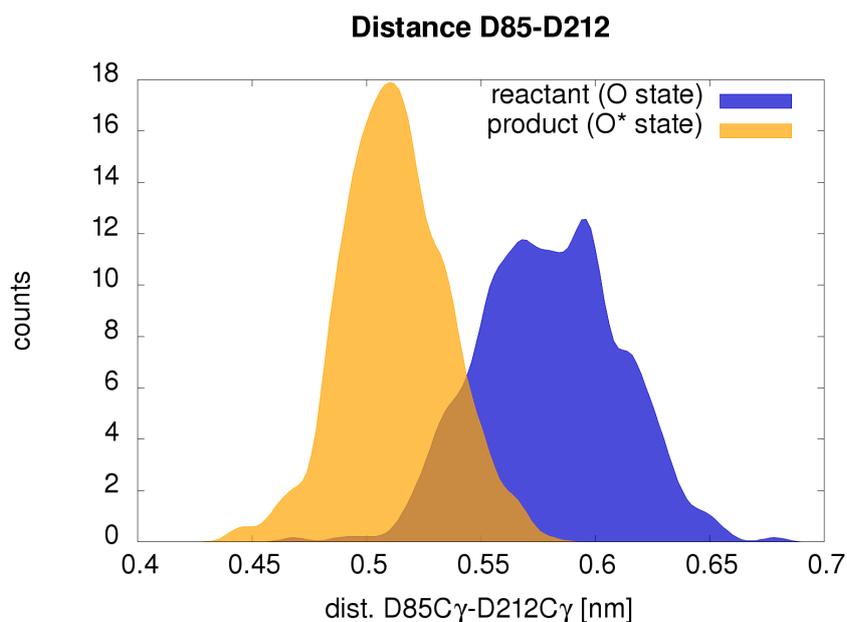


Figure 7.2: Histograms of the D85C γ -D212C γ distance for the O (blue area) and O* (yellow area) state. The population maximum for the O* state is reduced by approximately 0.1 nm compared to the O state.

population for each of the two reactant states. With respect to the population maxima, the D85C γ -D212C γ distance for the PT product is reduced by ca. 0.1 nm to 0.52 nm compared to the reactant state.

7.2 The Minimum Free Energy Path

In the previous section the most significant characteristics of the upper active site have been presented for the PT reactant (O) and product (O^*) state. In this section, there should be shed light on the PT reaction itself and the free energy change associated with it. Moreover, detailed information about the reaction mechanism should be given.

It has been shown that during the PT reaction from D85 to D212, the pentagonal HBN breaks down and as a result the two aspartates approach each other. In order to ensure convergence of the PT free energy profile, this parallel reaction has to be taken into account and also sampled. Accordingly, for the free energy calculation two reaction coordinates (two dimensions) will be used orthogonal to each other. In order to achieve this, the free energy profile for the PT from D85 to D212 will be calculated with 2D-MW-Metadynamics.

7.2.1 Simulation Setup

The starting structures of the MW-Metadynamics simulation are based on the last 100 ps QM/MM simulation of the reactant and product state (cf. section 7.1, p.63). Therefore, the geometries of the reactant and product trajectory were written out every 10 ps yielding to 20 structures in total. Each of these structures served as starting structure for one walker. For each walker, the same QM/MM setup, simulation parameters and water position restraints were used as for the simulations of the reactant and product state. Solely the simulation time was changed to 500 ps and the time step was adjusted to 0.5 fs. In order to describe the PT reaction from D85 to D212, the mCEC coordinate was used and defined in a Plumed input file. The mCEC definition contained all four water molecules of the QM/MM region and the D85/D212 oxygen atoms as possible proton coordination sites. The transfer of the excess charge was mapped relative to the D85C γ and D212C γ atom using the ζ_2 definition from equation 5.13. The PT coordinate was resolved with a σ value of 0.065. The second reaction coordinate aimed to sample the formation and destruction of the HBN. Since the description of this process is not straightforward, an indirect description based on the D85C γ -D212C γ distance was used. In order to gain a faster convergence of the free energy surface, the sampling space of this reaction coordinate was restricted to D85C γ -D212C γ distances below 0.66 nm. Above this threshold value a harmonic restraint with a spring constant of $\kappa = 200000$ kJ/(mol·nm²) set in. The distance dimension was resolved with a σ value of 0.02 nm. Moreover, a Gaussian hill size of $\omega_0 = 0.2$ kJ/mol, a deposition frequency of $\tau = 100$ fs and a bias factor of $\gamma = 8$ was applied. The free energy surface was created via post-processing of the HILLS files with the Plumed command *sum_hills*.

7.2.2 2D-MW-Metadynamics

Figure 7.3 depicts the two-dimensional free energy surface for the PT reaction from D85 to D212. The first dimension (x-axis) displays the free energy change along the ζ coordinate and hence the transfer of the proton (or excess charge), while the y-dimension represents the free energy change along the D85C γ -D212C γ distance which indirectly represents the HBN formation/destruction process. The whole free energy surface was sampled for 10 ns

in total (20 walkers, each 500 ps). This surface exhibits two minima basins: one for the reactant state at $\zeta = -0.5$ and one for the product state at $\zeta = 0.5$. Moreover, with respect to the D85C γ -D212C γ distance the reactant minimum is shifted to a larger value of around 0.57 nm compared to the product basin with 0.52 nm. The corresponding geometries and characteristics of these two states were already presented in the previous section in Figure 7.1 and Figure 7.2. In the reactant state, the pentagonal HBN is intact so that the D85C γ -D212C γ distance is increased, whereas it is the other way around for the product state. In terms of energetics, the reactant state represents the global minimum on the free energy surface and the product basin is raised by $\Delta G_{O-O^*} = 4.2$ kcal/mol. This means the proton is better stabilized on the D85 than on the D212. Moreover, with respect to these results, the PT from D85 to D212 and with this the O \rightarrow O* transition depicts an endergonic process. The minimum free energy path (① in Figure 7.3) which connects both basins, exhibits a barrier height of $\Delta G_{O-O^*}^\ddagger = 8.8$ kcal/mol. A transfer over this path is linked to a strong reduction of the D85C γ -D212C γ distance so that the approximate TS ($\zeta = 0.0$) exhibits an aspartate distance of 0.42 nm. A further analysis of the geometry of this state, reveals that the two aspartates D85 and D212 are directly facing each other and share the proton (see Figure 7.5, bottom, black dot). In other words, compared to the reactant and product state, in the TS there are no water molecules present between D85 and D212. This means that a PT over the minimum free energy path is direct, not water-mediated and connected with a complete breakdown of the HBN. Furthermore, the N ζ H group of the Schiff base lost with the absence of the water molecule its hydrogen bond acceptor in the TS so that this group interacts with one of the D212 carboxyl oxygens.

On the free energy surface in Figure 7.3 there is also another TS identifiable (labeled with ②), where the aspartate distance is increased to around 0.52 nm. Compared to the minimum free energy path, this pathway is raised by $\Delta G = 1.7$ kcal/mol. Interestingly, this TS features two different geometries. In one geometry the TS exhibits a hydronium (H₃O⁺) ion (see Figure 7.5, bottom, green dot), whereas the other transition geometry features a hydroxide (OH⁻) species (see Figure 7.5, bottom, red dot). The energetical differentiation between these higher lying water-mediated PT pathways and the corresponding transition geometries will be discussed in the next section.

7.3 Proton Transfer Pathways over Hydronium and Hydroxide Species

It has been shown that the PT reaction from D85 to D212 exhibits, beside the direct PT pathway, two other water-mediated pathways where the transfer is accomplished via hydronium and hydroxide species. In this section these two pathways will be further investigated. Particularly, the energetical differentiation of the two transfer mechanisms will be the central topic of this section. Moreover, the hydronium and hydroxide transition states will be further examined. In order to achieve this, a collective variable has to be found that is able to discriminate between the two PT pathways. This collective variable can then be used as additional, orthogonal reaction coordinate in a 3D-MW-Metadynamics simulation.

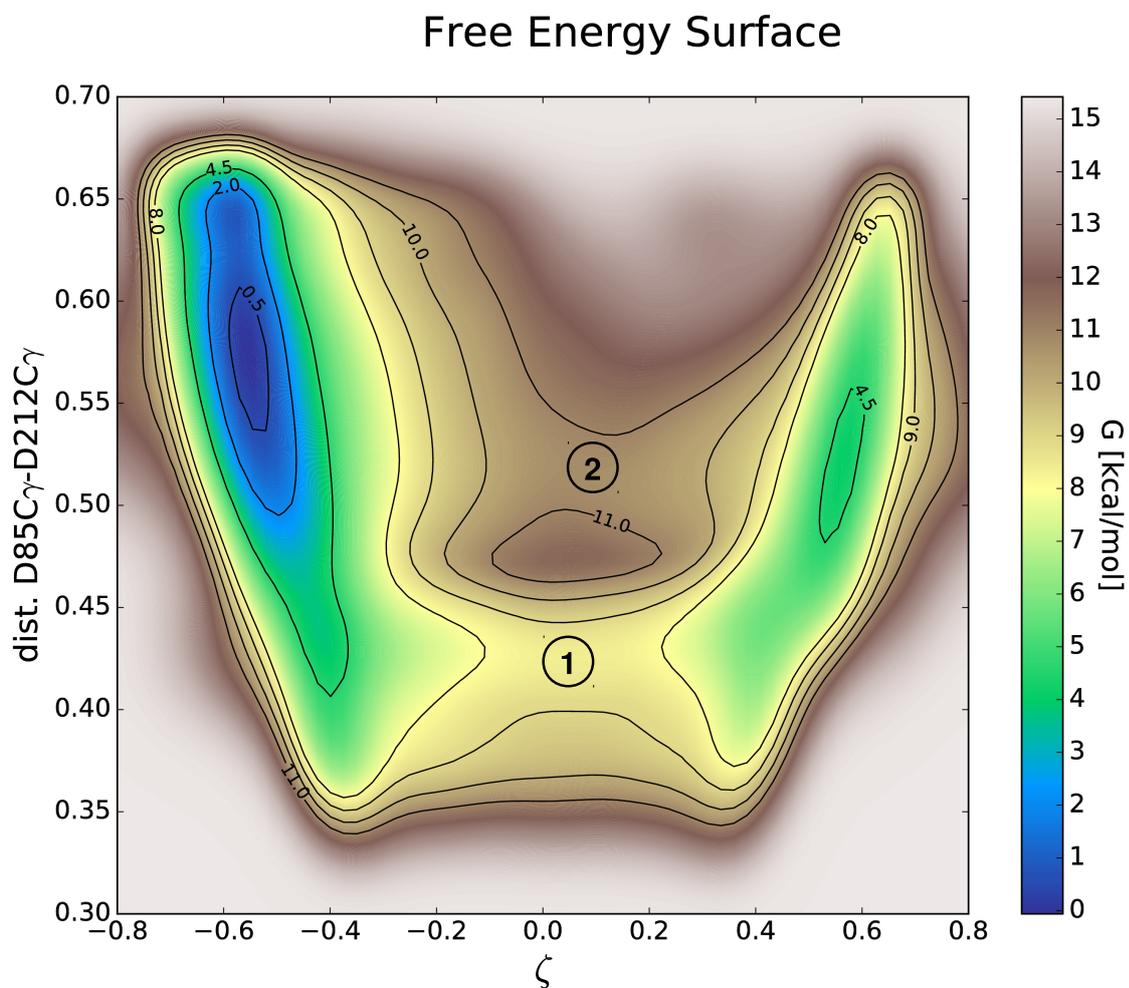


Figure 7.3: Gibbs free energy surface for the D85 \rightarrow D212 PT reaction depending on the D85C γ -D212C γ distance. The x-axis represents the relative transfer of the CEC between both amino acids. The y-axis describes via the D85C γ -D212C γ distance indirectly the HBN formation/destruction. There are two PT pathways identifiable labeled with ① and ②.

7.3.1 Distinction between Hydronium/Hydroxide Species and Simulation Setup

The two water-mediated pathways can be separated energetically with the Metadynamics approach. Therefore, a collective variable must be provided which is able to distinguish between the two transfer pathways. The most significant difference between both transitions is the occurrence of a hydronium or hydroxide ion. Hence, a plausible collective variable may be based on the hydrogen coordination number of the water oxygen atom. For a hydronium ion, the oxygen coordination number is then equal to three, whereas it is equal to one for a hydroxide ion. Since the QM region comprises four water molecules and each of them could occur in the PT as hydronium or hydroxide, the average water oxygen coordination number has to be used. For the description of the oxygen coordination number a Plumed-implemented function was used which reads:

$$s_O(r_{ij}) = \sum_{i \in O} \sum_{j \in H} \frac{1 - \left(\frac{r_{ij}}{r_0}\right)^n}{1 - \left(\frac{r_{ij}}{r_0}\right)^m} \quad (7.1)$$

The upper equation is a simple switching function which switches in a parameter-dependent region from one to zero. The parameters r_0 , n and m were manually fitted against water geometries containing hydronium or hydroxide species from the WATER27 benchmark set¹⁷⁴. A decent representation for the water oxygen coordination number yielded the parameter set with $r_0 = 1.2 \text{ \AA}$ and the exponents $n = 45$, $m = 90$. The functional behavior of $s_O(r_{ij})$ is depicted in Figure 7.4A. According to that, a hydrogen atom is treated as coordinated/bound to the water oxygen atom if its distance is less than 1.1 \AA . Contrary to this, the hydrogen atom is seen as decoordinated/unbound if the r_{ij} distance is greater than 1.3 \AA . In Figure 7.4B the coordination function was applied to four different hydronium/hydroxide-containing systems of the WATER27 set. For the hydronium ion we get a maximum water oxygen coordination of $s_O(r_{ij}) = 3.0$. It is the other way around for the hydroxide ion, where $s_O(r_{ij})$ displays the minimum water oxygen coordination of 1.0. In order to evaluate cluster systems like the Zundel ion ($\text{H}_3\text{O}^+\text{H}_2\text{O}$) or $\text{OH}^-(\text{H}_2\text{O})_3$, the average of the water oxygen coordination function $\overline{s_O(r_{ij})}$ must be used. Generally, for water clusters which exhibit a hydronium ion, $\overline{s_O(r_{ij})}$ will be greater than 2.0. In contrast to this, $\overline{s_O(r_{ij})}$ will display a value below 2.0 for water clusters which contain a hydroxide ion. The Zundel ion in Figure 7.4B displays a special case where the coordination function also provides a decent result. The proton/excess charge is shared equally between the two water molecules in this system. Hence, the $s_O(r_{ij})$ function is situated between the mentioned extreme cases in the region of $r_{ij} = 1.2 \text{ \AA}$.

The previous results have shown that a proper collective variable was found that is efficient enough to distinguish between the conventional Grotthuss mechanism (over hydronium) and a proton hole (over hydroxide) pathway. In order to separate both pathways energetically, the in subsection 7.2.1, p.66 presented 2D-MW-Metadynamics setup was used and solely extended by $\overline{s_O(r_{ij})}$ as a third reaction coordinate. The $\overline{s_O(r_{ij})}$ collective variable was defined such that it represents the average water oxygen coordination number of all four

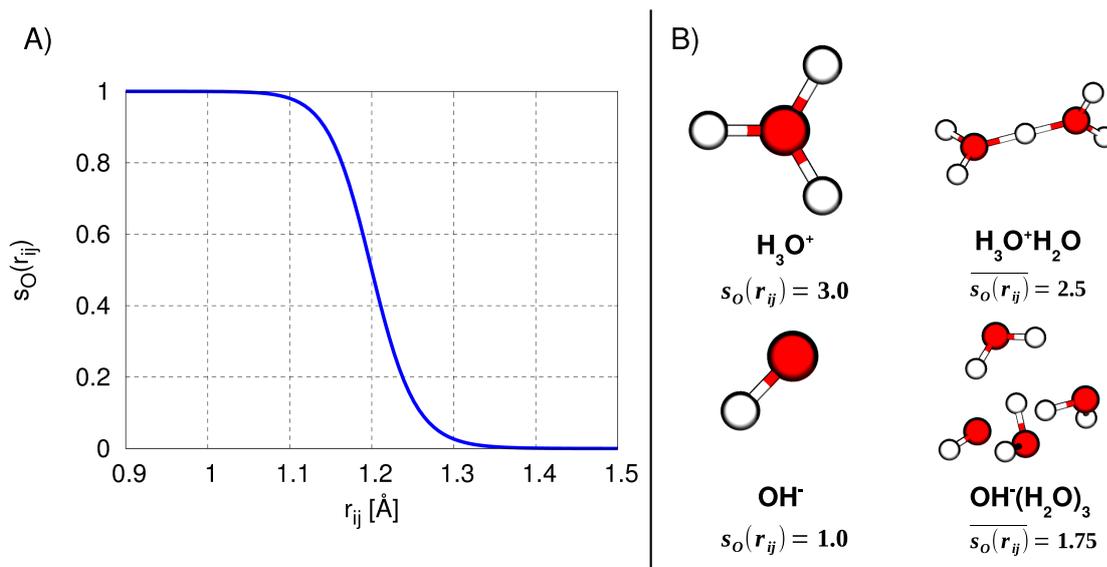


Figure 7.4: A: The coordination function (equation 7.1) with the following parameters: $r_0 = 1.2 \text{ \AA}$, $n = 45$, $m = 90$ is plotted over a r_{OH} range of $0.9 - 1.5 \text{ \AA}$. B: Different hydronium- and hydroxide-containing systems are shown together with their corresponding $s_O(r_{ij})/\overline{s_O(r_{ij})}$ function. If a hydronium ion is present the coordination function is greater than 2.0, if a hydroxide ion is present the coordination function is less than 2.0.

water molecules present in the QM region. The Metadynamics resolution of $\overline{s_O(r_{ij})}$ was set to $\sigma = 0.025$. The complete Plumed input file for the 3D-MW-Metadynamics simulation is attached in the Appendix (see Listing A.1). Again, the Plumed `sum_hills` command was used to yield a $\overline{s_O(r_{ij})}$ vs. ζ projection of the free energy surface.

7.3.2 3D-MW-Metadynamics

Figure 7.5 depicts the PT with respect to the transfer pathway for the D85 to D212 reaction. The x-axis represents, like in Figure 7.3, the free energy change along the CEC transfer, while the y-dimension displays the free energy change with respect to the average water oxygen coordination number of the four QM-treated water molecules. The free energy surface exhibits two minima at $\zeta = \pm 0.5$. Like in Figure 7.3, these basins represent the PT reactant and the energetical higher lying product state which were already discussed in section 7.1, p.63. Furthermore, there are three possible pathways identifiable which can accomplish a PT between these basins. The minimum free energy path on this surface leads over an average water oxygen coordination number of $\overline{s_O(r_{ij})} = 2.0$. Thus, the $\overline{s_O(r_{ij})}$ coordinate reveals that no hydronium or hydroxide species are occurring in this pathway and it must be a direct, non-water-mediated PT. This pathway is already known and was further examined in section 7.2, p.66: it is the direct PT from D85 to D212 with a reaction barrier of $\Delta G_{O-O^*}^\ddagger = 8.8 \text{ kcal/mol}$.

The second lowest free energy path represents the one which exhibits an average coordination of $\overline{s_O(r_{ij})} = 1.75$. The maximal barrier height for this PT pathway amounts to

ca. $\Delta G^\ddagger = 11.6$ kcal/mol. The corresponding TS structure (see Figure 7.5, bottom, red dot) shows, like expected for a $\overline{s_O(r_{ij})}$ value below 2.0, the presence of a hydroxide ion in the water chain. The key characteristics of this TS are the two aspartates D85 and D212 which are protonated. Furthermore, the water-mediated nature of this PT mechanism is noticeable: While the direct PT (cf. Figure 7.5, bottom, black dot) completely destroys the HBN, it is only perturbed for the proton hole pathway and rather serves with its hydrogen bonds as a stabilizing scaffold for the hydroxide ion. Furthermore, the D85-D212 distance is so large that the charge transfer via the proton hole is accomplished over two water molecules.

The third PT pathway shows an average coordination of $\overline{s_O(r_{ij})} = 2.25$. This pathway exhibits with ca. $\Delta G^\ddagger = 11.9$ kcal/mol nearly the same barrier height as the proton hole pathway. Since $\overline{s_O(r_{ij})}$ is greater than 2.0 the corresponding TS contains a hydronium ion. The corresponding geometry is shown in Figure 7.5, bottom, green dot. Contrary to the proton hole pathway, in this TS, both aspartates D85 and D212 are deprotonated. Nevertheless, similar to the proton hole pathway, also this process is water-mediated and the HBN is only perturbed rather than destroyed. Furthermore, this TS gets stabilized by the hydrogen bonds which were formed by the two aspartates and the neighboring water molecules. Contrary to the proton hole pathway, the hydronium one is accomplished over just one water molecule.

Beside the lower barrier height, the in Figure 7.5 presented free energy surface yields another hint considering the preference for the proton hole over the hydronium pathway: The formation of the proton hole, or to put it another way, the protonation of the D212 is with around $\Delta G = 11.0$ kcal/mol even more favorable than the formation of the hydronium ion. This is an interesting fact because from chemical intuition, one would assume that the aspartate side chain with its low pK_a of 3.9¹⁷⁵ is rather deprotonated than the other way around. However, the aspartate pK_a value was determined in aqueous solution with a dielectric constant of $\epsilon = 80$ ¹⁷⁶ but the electrostatics in a protein environment are totally different^c. Thus, as a sufficient condition the semiempirical QM method DFTB3 was benchmarked regarding its preference for the hydronium/proton hole pathway against a higher level DFT method.

7.4 A DFTB3 Benchmark Study: The AspH⁺-H₂O-Asp System

In the previous section, it has been shown that DFTB3 prefers – counterintuitively – the proton hole pathway over the hydronium one for the O→O* PT reaction. Moreover, both pathways exhibit nearly the same reaction barrier on the free energy surface in Figure 7.5. Thus, the question if DFTB3 suffers from a preference for the proton hole pathway is a qualified one and will be answered in this section. In order to achieve this, the D85→D212 PT system is simplified to a water-bridged, protonated aspartate-aspartate system (see Figure 7.6, bottom). Based on this toy system, the energetics of the PT and its pathway

^cFor protein interior: $\epsilon \approx 6 - 7$, for protein surface: $\epsilon \approx 20 - 30$ ¹⁷⁶.

Free Energy Surface

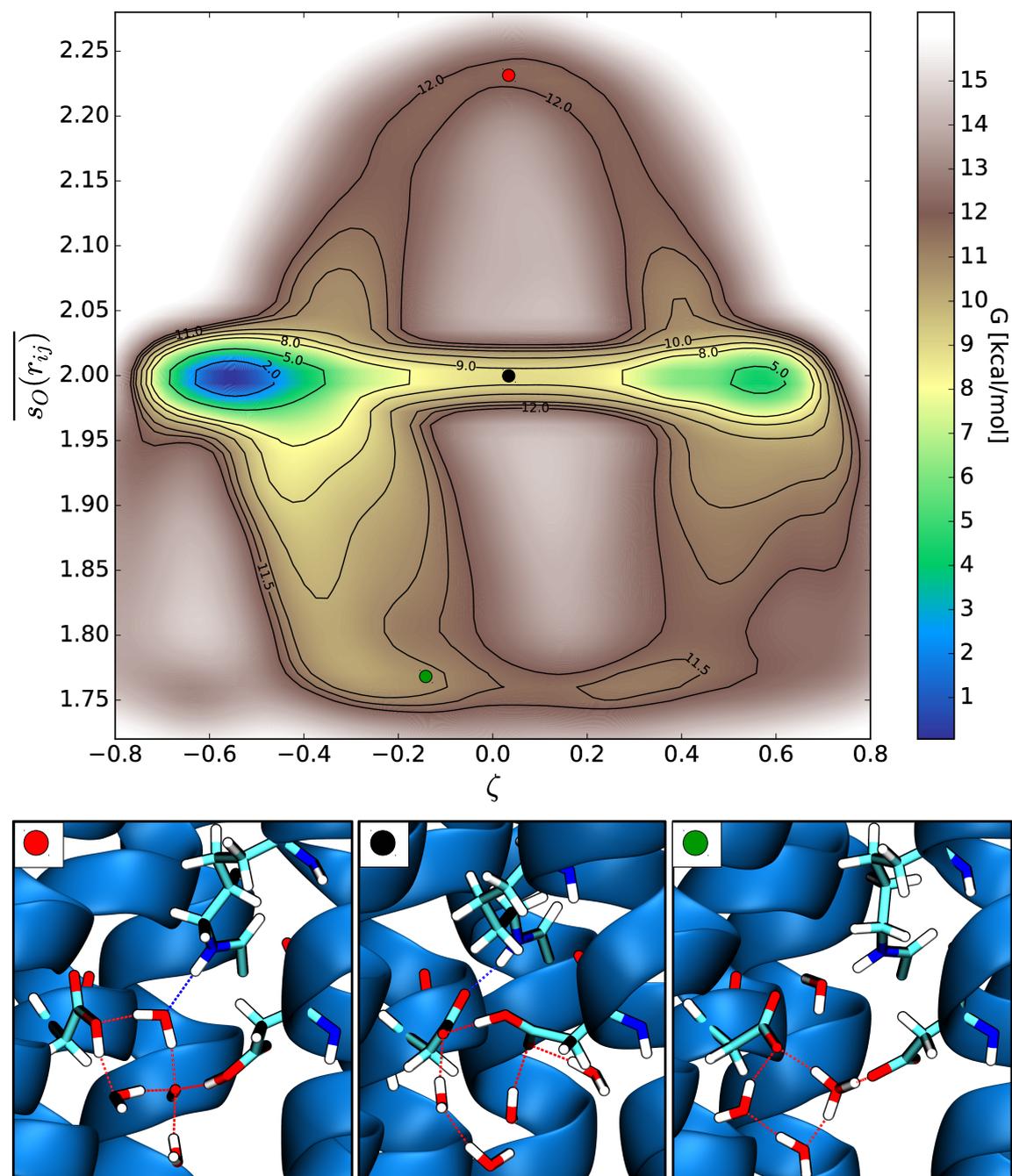


Figure 7.5: Top: Gibbs free energy surface for the D85→D212 PT reaction depending on the $s_O(r_{ij})$ function. The x-axis represents the relative transfer of the CEC between both amino acids. The y-axis reflects the average water oxygen coordination number of the QM-treated water molecules. There are three PT pathways identifiable which differ in their barrier heights. Furthermore, their transition structures are labeled by a red, black and green dot respectively. Bottom: The upper active site geometries of the marked PT TS structures are displayed. The geometry which belongs to the minimum free energy path (black dot) shows a direct D85-D212 interaction, while the geometries of the higher lying pathways exhibit hydroxide (red dot) or hydronium (green dot) species.

will be analyzed, employing 2D-MW-Metadynamics with DFTB3 and the hybrid DFT method B3LYP.

7.4.1 Simulation Setup

For the model setup a reactant state geometry from section 7.1 was used. Two-dimensional Metadynamics simulations are very expensive with a DFT method like B3LYP. Hence, the reactant system was simplified with the objective to design a smallest possible QM region that still features the key characteristics of the D85→D212 PT system. The resulting system comprised a QM region^d of 18 atoms. It contained the protonated aspartate side chain, resembling D85, which is bridged over one water molecule to a deprotonated aspartate side chain, which should represent D212. Moreover, in a second geometry, the protonation states of the aspartate side chains were inverted in order to model the PT product state.

The DFTB3 and B3LYP benchmark calculations were set up and conducted equally. For the DFTB3 calculations, again Gromacs v.5.0 with the implemented DFTB3 code in combination with the 3OBw parameter set and the D3BJ dispersion correction was used. In order to conduct the DFT calculations, Gromacs v.5.0 was interfaced with the QM package ORCA v.3.0.1¹⁷⁷. The hybrid density functional B3LYP^{107,109–111} was used. Since Kaila et al.¹⁷⁸ suggested for a correct description of PT energetics a triple- ζ basis set, the def2-TZVP basis set^{179,180} was used in combination with the D3BJ dispersion correction. Furthermore, in order to accelerate the SCF calculations, the RIJCOSX^{181,182} approximation was used^e.

First, the geometries of the modeled reactant and product PT toy systems were minimized with the steepest descent algorithm. For the 2D-MW-Metadynamics simulations the QM-interfaced Gromacs versions were used in combination with the Plumed v.2.1.1 plugin. The simulations were conducted at 300 K in vacuo using the stochastic dynamics integrator in Gromacs. In order to prevent PT disturbing, dynamical effects like rotation, spatial separation etc. of the two aspartates, the carbon atoms of the side chains were fixed with the mdp-file option *freeze-grps*.

Two Metadynamics collective variables were defined: The first reaction coordinate describes the PT reaction between both aspartates. The mCEC definition (ξ) contained all transfer protons of the water molecule and aspartate residues. The oxygen atoms of the aspartate carboxyl groups and the water molecule were taken into account as possible proton coordination sites. With respect to the fact that all oxygen atoms of the carboxylic groups could act as a proton donor (D1, D2) or acceptor (A1, A2), we modified the relative ζ_1 representation in equation 5.13 to:

$$\zeta = \frac{d_{\xi,D1}d_{\xi,D2}}{d_{\xi,D1}d_{\xi,D2} + d_{\xi,A1}d_{\xi,A2}} \quad (7.2)$$

The second collective variable represents with $s_O(r_{ij})$ the oxygen coordination number of the bridging water molecule. The Metadynamics simulations consisted of 16 walkers respectively. Eight walkers started in the reactant state and the other eight in the product state.

^dNo MM part.

^eThe speedup effect is due to two reasons: The Coulomb term J is treated via the resolution of identity approximation (RI)^{183–185} and the Exchange term X via seminumerical integration.

Each of them had a simulation time of 12.5 ps. Hence, in total the PES was sampled for 200 ps. For the general Metadynamics parameters, a Gaussian hill size of $\omega_0 = 0.3$ kJ/mol, a deposition frequency of $\tau = 5$ fs and a bias factor of $\gamma = 4$ was used. The resolution of the ζ dimension was set to $\sigma = 0.067$, while for the $s_O(r_{ij})$ dimension a σ of 0.133 was used.

7.4.2 Comparison of DFTB3 with B3LYP/def2-TZVP

The $s_O(r_{ij})$ vs. ζ PES for DFTB3 and B3LYP are depicted in Figure 7.6. The reactant and product state conformations form minima on these surfaces at $\zeta = 0.0$ and $\zeta = 1.0$ respectively. In both methods these minimum conformations are isoenergetic in potential energy, which means the proton gets equally stabilized on both aspartates. The minimum energy path for the PT is accomplished on both surfaces over a $s_O(r_{ij})$ value of 1.0 and hence over the proton hole mechanism. For DFTB3 the barrier height for this reaction amounts to ca. $\Delta E = 4.3$ kcal/mol, whereas in the B3LYP representation this barrier is raised by 1.0 kcal/mol to $\Delta E = 5.3$ kcal/mol. In contrast to the results for the analogous PT in the protein environment (cf. section 7.3, p.67), the hydronium pathway lies for both methods a couple of kcal/mol higher and hence is energetically more separated from the proton hole path. In the B3LYP reference calculation, the hydronium path exhibits a barrier height of $\Delta E = 7.4$ kcal/mol and lies 2.1 kcal/mol higher than the proton hole path. Contrary to this, the DFTB3 results doesn't follow the observed underestimation compared to B3LYP for this pathway: The hydronium path shows a barrier height of ca. $\Delta E = 8.8$ kcal/mol and lies with this 4.5 kcal/mol higher than the proton hole pathway.

Let us have a closer look on the TS structures in order to answer the question, why the proton hole pathway is energetically more favorable than the hydronium one. The TS for the proton hole path is displayed in Figure 7.6, bottom, red dot. In this conformation, the oxygen atom of the proton hole is twice stabilized by the hydrogen bonds from the two flanking, protonated aspartates. Furthermore, the excess charge is centered in the middle of this hydrogen-bonded complex. The case is different for the hydronium TS. This conformation also exhibits two hydrogen bonds. Nevertheless, these are both donated by the hydronium ion and hence not so strong. Furthermore, the excess charge is more located on both flanking aspartates and hence more difficult to stabilize in the gas phase.

With respect to the time exposure of the calculations, DFTB3 definitely outperforms B3LYP: The $s_O(r_{ij})$ vs. ζ free energy surface was obtained with B3LYP in 68 days and 19 hours. In contrast, DFTB3 needed for this calculation only five minutes. The presented results show that the efficient DFTB3 method is able to predict the correct PT pathway and furthermore yields a reasonable barrier height which is in accordance with the expensive DFT reference. Nevertheless, for the hydronium pathway the barrier height is increased with DFTB3 and also overestimates the B3LYP results. For PT reactions like in section 7.3, where the proton hole and hydronium pathways lie energetically very close to each other, this could pose a problem to DFTB3 with respect to the correct prediction of the pathway.

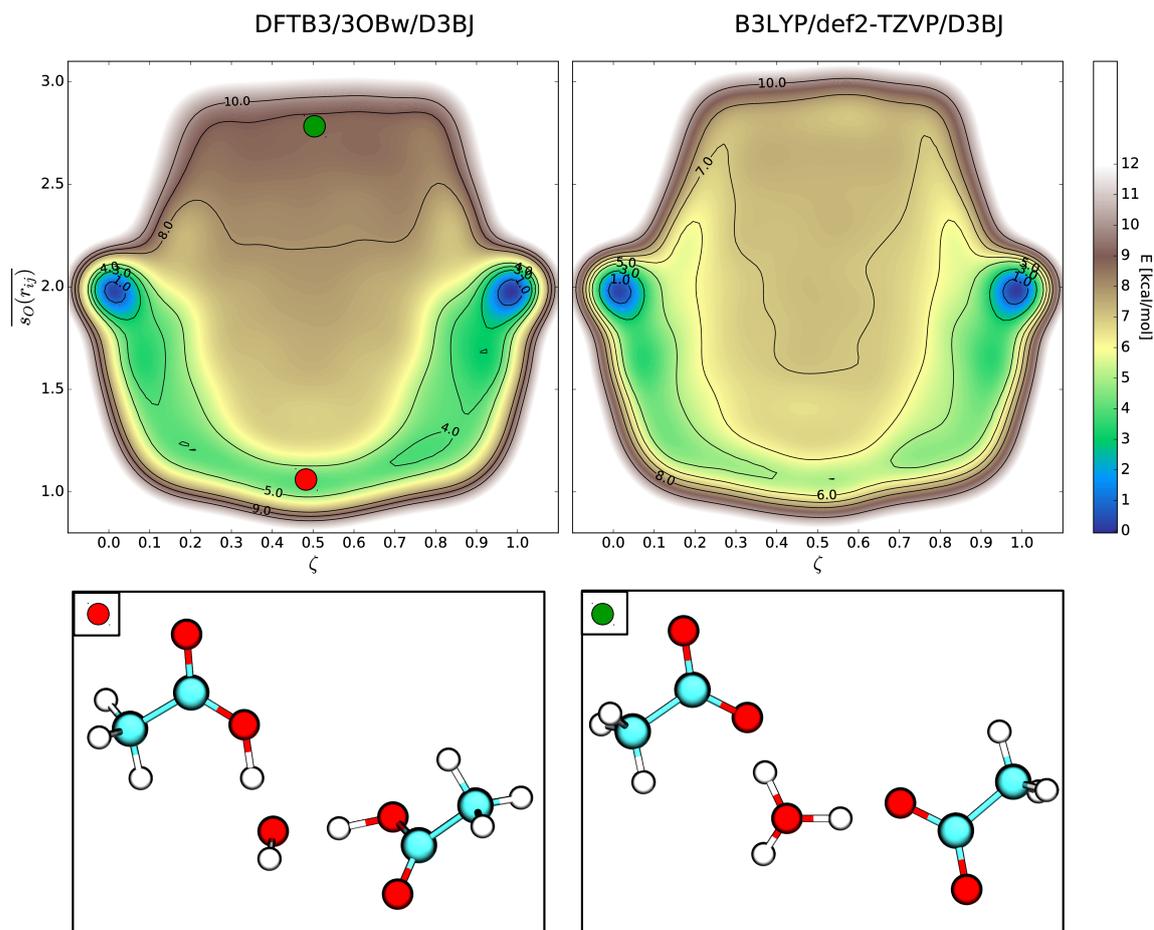


Figure 7.6: Top: PES for the PT reaction in the $\text{AspH}^+\text{-H}_2\text{O-Asp}$ system calculated with DFTB3/3OBw/D3BJ (left) and B3LYP/def2-TZVP/D3BJ (right). The x-axis of the contour plots represents the transfer of the CEC between the two aspartates. The y-axis indicates the coordination number of the bridging water molecule. Both methods show a preference for the proton hole pathway. Bottom: DFTB3 TS structures of the proton hole (red dot) and hydronium pathway (green dot).

7.5 Conclusions

In this chapter the $\text{O} \rightarrow \text{O}^*$ state transition and with this the PT reaction from D85 to D212 has been investigated. The free energy change which is connected with the PT was resolved by employing DFTB3-based QM/MM simulations in combination with multidimensional MW-Metadynamics. Furthermore, structural characteristics of the PT reactant, product and transition states were investigated.

With respect to the Metadynamics simulations, the minimum free energy pathway of the $\text{D85} \rightarrow \text{D212}$ PT reaction exhibits a barrier height of $\Delta G_{\text{O-O}^*}^\ddagger = 8.8$ kcal/mol. Consequently, this process represents a kinetically feasible reaction. Regarding the PT mechanism, the minimum free energy pathway features a direct charge transfer between D85 and D212. Beside this mechanism, two additional, higher lying pathways were resolved and identified.

In these, the PT is water-mediated and their TS exhibit either hydronium or hydroxide-like species. The use of the average water oxygen coordination number served as an appropriate collective variable in order to discriminate energetically between the hydronium and proton hole pathway. Since these PT reaction pathways lie approximately 3 kcal/mol higher than the direct one, it can be assumed that the PT from D85 to D212 is accomplished via a direct PT rather than via a water-mediated (hydronium/proton hole) one. It is worth mentioning that the energetics of the direct and proton hole pathway were not discussed in literature until now. Only the hydronium pathway was considered by Phatak et al.^{67f} and is in accordance with the here presented results.

The comparison of the reaction barriers of the hydronium and proton hole pathway revealed that the proton hole mechanism is, with respect to energetics, slightly more favorable. In order to evaluate the reliability of DFTB3 with respect to the description of PT reactions in general and specifically for the discrimination of the hydronium and proton hole pathway, we benchmarked the semiempirical method against the computationally more expensive B3LYP-DFT method. DFTB3 yielded in these benchmark calculations reasonable results with a small reaction barrier underestimation and a slight preference for the proton hole pathway. Nevertheless, we assume that the preference for the proton hole over the hydronium pathway in bR results from the electrostatic environment of the protein.

With a ΔG_{O-O^*} of 4.2 kcal/mol, the D85 \rightarrow D212 PT represents an endergonic reaction. According to this, the $O \rightarrow O^*$ state transition can not exist as an autonomous reaction. It is rather the case that the D212 serves as an intermediate proton carrier in the $O \rightarrow$ bR transition which was already proposed by Dioumaev et al.⁶⁵ and Phatak et al.⁶⁷.

Furthermore, it has been shown that the PT from D85 to D212 also affects the environment around the two amino acids. The most significant change is the disruption (hydronium/proton hole pathway) or whole break down (direct pathway) of the pentagonal HBN during the PT. Hence, it can be assumed that also the $O \rightarrow$ bR transition is accompanied with rearrangements in the pentagonal HBN.

^f $\Delta E_{O-O^*} \approx 12$ kcal/mol, based on CPR calculations.

O→bR: A Long-Range Proton Transfer

The last chapter has shown that the O* state serves as an intermediate state for the assumed long-range PT reaction from D85 to the PRG group.

This chapter tries to close the gap between the O* state and the bR ground state by simulating the PT from D212 down to the extracellular lying PRG. Or, to put it another way, the bR state will be recovered by starting from an O* state structure. A prerequisite for this long-range PT – the presence of a water wire which connects the proton donor and acceptor – was already found in chapter 6.

The free energy profile for the O*→bR transition will be in the focus and is resolved via Umbrella Sampling. The prediction of free energies for this transition is a delicate issue, since beside the PT all other side processes that are taking place have to be equilibrated in order to obtain a converged free energy profile. To these mentioned side processes mainly belong the formation of the shared proton conformation in the PRG, the upswing movement of the R82 side chain and the regeneration of the HBN.

In the previous chapter it was found out that the upper active site shows for the PT process between D85 and D212 a slight preference for the proton hole pathway. Hence, another interesting topic which will be analyzed is the pathway that the PT takes from D212 down to the PRG.

Finally, a complete reaction mechanism will be proposed for the O→bR transition. This assumption will then be compared with experimental results that are known for this process.

8.1 Simulation Setup

As starting structure for the simulations of the long-range PT served a geometry from the O* state HREX simulation close to 5.3 ns (see Figure 6.9C, p.62). This conformation exhibited a continuous water wire between D85/D212 and the PRG region. The structure was converted to a QM/MM system in which the QM region comprised the D85, D212, R82, E194 and E204 side chains^a. Furthermore, eight water molecules that were present in the

^aFrom C β on.

region between D85/D212 and the PRG were also treated quantum mechanically. The QM region contained 78 atoms in total. Again, this region was treated with DFTB3/3OBw with D3BJ dispersion correction. The rest of the system was described with classical mechanics using the CHARMM36 force field. The simulations were conducted with the DFTB3-implemented Gromacs v.5.0. For the handling of the ζ coordinate and the average water oxygen coordination number the Plumed plugin v.2.1.1 was used. The mCEC coordinate ξ was defined such that all carboxyl oxygens of D85, D212, E204 and all QM-treated water oxygens could serve as possible proton coordination sites. All protons of the QM-treated water molecules and the D212 proton were taken into account to be transferred. The relative mapping of the CEC was accomplished via a ζ definition like in equation 5.13, p.43 (ζ_2). As initial donor the D85C γ atom and as final acceptor the E204C δ atom was chosen. This definition ensures that both carboxyl oxygens could serve as donor or acceptor. For the definition of the $\overline{s_O(r_{ij})}$ function all oxygens and hydrogens of the shuttling water molecules and the D212 proton were taken into account. First, the starting structure was modified such that the D212 side chain was deprotonated and a water molecule of the water wire was protonated. The resulting structure exhibited a ζ value of 0.08 and a $\overline{s_O(r_{ij})}$ value of 2.13. This modified structure was then geometry-optimized with the steepest descent algorithm. Then, a harmonic restraint with a spring constant of $\kappa = 4000$ kJ/mol was placed on the ζ variable, centered at $\zeta = 0.3$, such that the CEC was directly located on the water wire between D212 and the PRG. A restrained NVT and NPT equilibration followed where the ζ restraint was kept on the mentioned value. The QM/MM system was shortly pre-equilibrated for 500 fs at 300 K in the NVT and for 1 ps in the NPT ensemble at 300 K and 1.0 bar. For the temperature and pressure coupling, the Nosé-Hoover thermostat and the Parrinello-Rahman barostat were used. The applied time step was 0.5 fs. Moreover, in order to prevent the leakage of QM-treated H₂O molecules out of the cavity, spherical harmonic restraints were applied to the water oxygen atoms with a radius of 0.4 Å and a spring constant of $\kappa = 2000$ kJ/mol. The NPT equilibrated structure was taken as starting structure for 40 Umbrella Sampling simulations. These were conducted in the NPT ensemble with a time step of 0.5 fs at 300 K and 1.0 bar. The spherical harmonic restraints on the QM-treated water molecules were loosened by applying the same spring constant as before but a radius of 2.5 Å. In each window the starting mCEC position of $\zeta_s = 0.3$ was changed via a time-dependent harmonic restraint ($\kappa = 4000$ kJ/mol) over a period of 3 ps to a window-specific end value ζ_e . The ζ_e values were chosen in equidistant steps ranging from -0.44 (proton localized on D212) to 0.8 (proton localized on E204). After 3 ps the harmonic restraint became time-independent (stationary) and the window was simulated until 800 ps. The first 500 ps were seen as equilibration time. Hence, the data for the analysis of the Umbrella sampling windows and for PMF calculation was based on the last 300 ps of simulation.

8.2 Umbrella Sampling for the $O^* \rightarrow bR$ Transition

Figure 8.1A shows the ζ -restrained geometry-optimized structure where the D212 proton was directly placed on the water wire between D212 and E204. The ζ value for this conformation is 0.29 and the hydronium ion is approximately at the same height as the positively charged R82 side chain. During the short NVT pre-equilibration of 500 fs, an interesting finding could be made: The water wire lost its hydronium ion and instead exhibits a proton hole, a protonated D212 and a protonated E204. With respect to Figure A.5 the pathway change happens very early after around 150 fs. Moreover, this change is not indicated by the ζ variable, but still it is reflected by $\overline{s_O(r_{ij})}$, which was introduced for this purpose. This observation of the pathway change is not unique to the system setup. The equilibration from the hydronium path into the proton hole path was also observed for other QM/MM setups of this PT system (not shown here). The proton hole formation happened there in most Umbrella Sampling runs. Furthermore, a back conversion from the proton hole into the hydronium pathway was never observed. Thus, it can be assumed that the D212 \rightarrow E204 PT shows, like the D85 \rightarrow D212 PT, a preference for the proton hole path over the hydronium path. A possible explanation therefore could be the electrostatic environment (like discussed before) but also the presence of the R82 side chain close to the PT pathway. The latter assumption is justified by the fact that during the hydronium-based PT, a positive charge passes by the positively charged R82 side chain, whereas in the proton hole-based pathway it is a negative charge that passes by. The visual analysis of the Umbrella Sampling simulations reveals a special event for a few simulations where the proton hole is spatially close to the R82 side chain (region from $\zeta = -0.3 - 0.1$): In these windows the proton hole gets reprotonated from one of the amine groups of the R82 guanidinium group. As a result, this event simply aborts the PT process. Since it is well known that the error of nitrogen proton affinities is striking in DFTB⁸¹, we believe that the deprotonation of arginine is an artifact. Nevertheless, these windows which suffer from this inappropriate description, have to be found and must be excluded from the subsequent WHAM analysis. In order to accomplish this, the last 300 ps of the Umbrella Sampling simulations were post-analyzed with the already introduced coordination function $\overline{s_O(r_{ij})}$. This time, the average hydrogen coordination number of the guanidinium nitrogen atoms $N\epsilon$, $N\eta_1$, $N\eta_2$ was used. With this function the artificial deprotonation of the guanidinium group can be identified when a window exhibits a $\overline{s_N(r_{ij})}$ value of 1.33. On the other hand, for windows where the proton hole mechanism is still intact, this function evaluates to 1.66. During this analysis, four windows were identified where the guanidinium group got deprotonated. Consequently, these windows were not taken into account for the subsequent WHAM analysis. For the remaining 36 windows, the configuration of the water wire was analyzed with the $\overline{s_O(r_{ij})}$ function: All windows either were in the proton hole (transition) conformation or exhibited a normal water wire (reactant/product states). Since the windows showed a reasonable overlap on the ζ reaction coordinate (see Figure A.7) they were used for a WHAM analysis.

The calculated PMF profile is depicted in Figure 8.2. The basin for the PT reactant state is localized around $\zeta = -0.33$, while the PT product basin is centered around $\zeta = 0.82$. The PMF indicates that the PT product lies with $\Delta G = -5.4$ kcal/mol, much lower than the

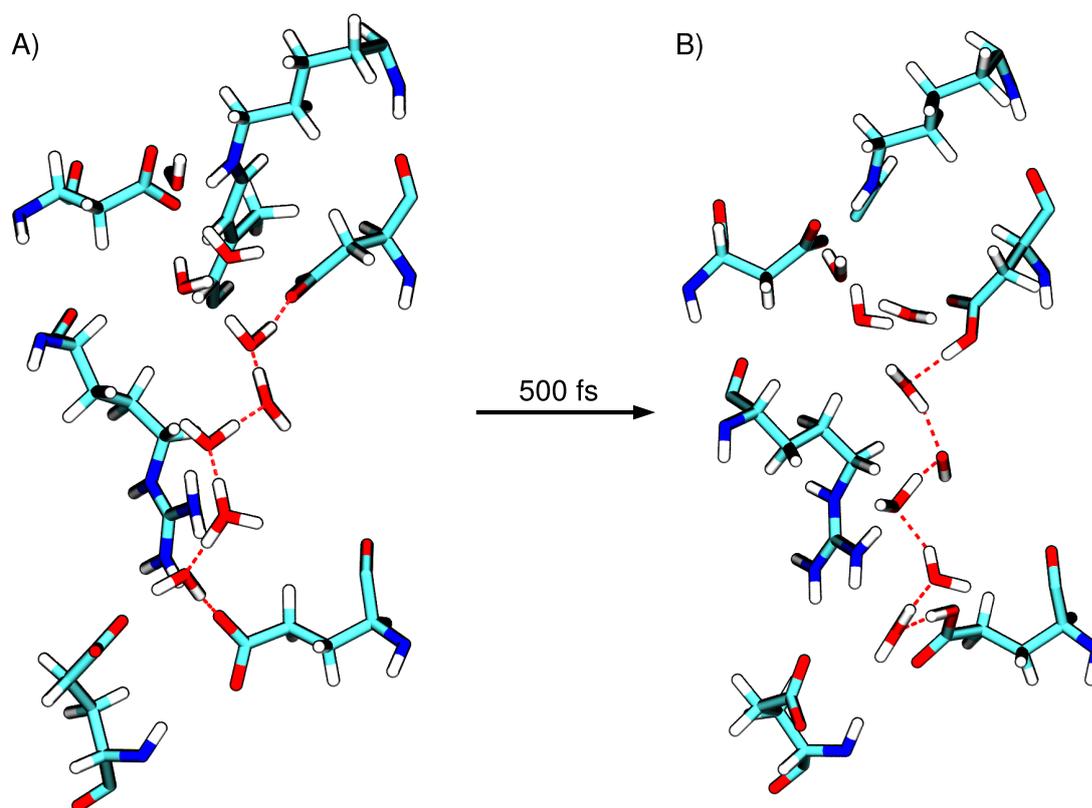


Figure 8.1: A: Geometry-optimized O^* state structure showing the complete active site together with a conducting water wire. The H^+ has been directly placed on the water wire. B: After 500 fs NVT equilibration the CEC is still centered between D212 and E204 but the two amino acids are protonated and the water wire contains a proton hole now.

PT reactant on the free energy profile. The rate-limiting step of this reaction can be found at $\zeta = 0.20$ and shows a barrier height of $\Delta G^\ddagger = 4.7$ kcal/mol. Regarding these values, the PT reaction from D212 \rightarrow E204 represents an exergonic and hence spontaneous process. The free energy barrier associated with the rate-limiting step suggests that a barrier crossing at room temperature is feasible and furthermore justifies the assumption that the O^* state serves only as a metastable intermediate in the $O \rightarrow bR$ transition. Consequently, the proton can be easily transferred down to the PRG, if a conducting water wire has been formed. In order to further understand this PT process, one has to analyze the molecular structures corresponding to minima and maxima on the PMF profile. The reactant state (orange dot in Figure 8.2), is very similar to the MM-based starting geometry (Figure 6.9C, p.62) which structural properties have already been discussed. Figure 8.2, green dot depicts the structure which corresponds to the TS at $\zeta = 0.2$. In this geometry, the proton hole is located half-way between the protonated D212/E204. The proton hole is stabilized by the surrounding water molecules. Compared to the reactant geometry, the guanidinium group of R82 reoriented towards E194 in order to form hydrogen bonds with this amino acid. This

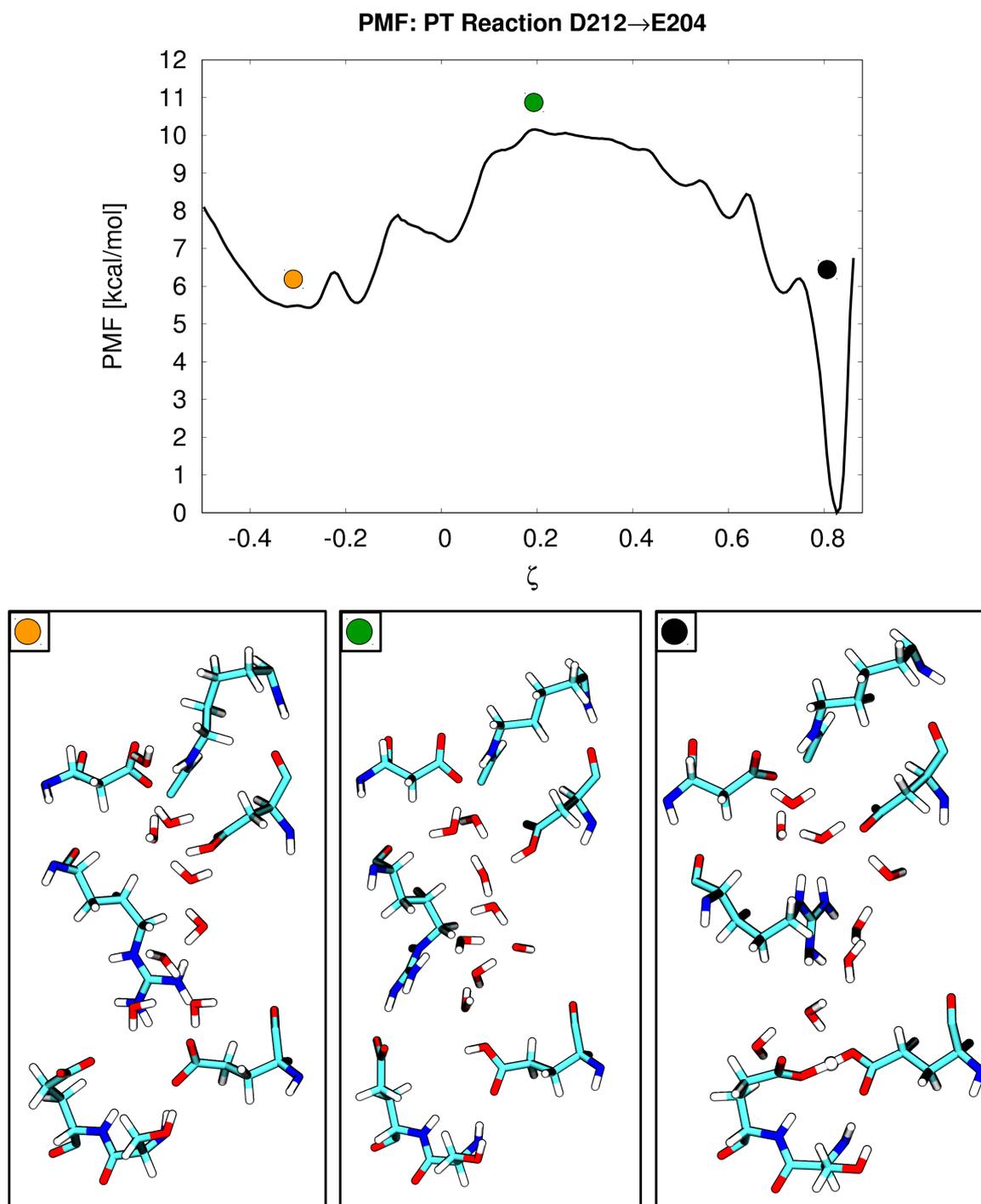


Figure 8.2: Top: PMF for the PT reaction from D212→E204 over the proton hole pathway. Bottom: The geometry of the reactant (orange dot), product (black dot) and an intermediate state (green dot) are shown. The QM/MM sampled reactant state is comparable to the structure shown in Figure 6.9C. In the TS geometry D212 and E204 are both protonated and the proton hole is stabilized by surrounding water molecules. The geometry of the product state shows that the shared proton conformation of the PRG has been formed during simulation. Furthermore, the R82 side chain is reoriented towards the upper active site.

reorientation results probably from the protonation of E204, which makes the hydrogen-bonding situation for the guanidinium group unfavorable. With respect to the product state structure (black dot in Figure 8.2), two important observations can be made: Firstly, the ground state PRG conformation with a shared proton predicted by Goyal et al.⁴⁹, has been formed spontaneously during simulation. The proton is shared between E194 and E204, while S193 is stabilizing this complex over hydrogen bonds. Secondly, the R82 side chain reoriented during simulation towards the upper active site and is therewith in its swung-up conformation like it is assumed for the bR state. This reorientation is again a result of the hydrogen bonding situation. With the formation of the PRG, the R82 side chain simply lost its hydrogen bonding partners in the swung-down conformation. However, the reformation of the HBN is another structurally important property of the bR state, which has not yet been recovered during simulation. Instead, the HBN is still disrupted and the retinal N ζ H group hydrogen bonds with D85. Another interesting point on the PMF curve, whose corresponding structure is shown in Figure A.8, is $\zeta = 0$. The minimum in the PMF results from a geometry where the proton hole is stabilized (not protonated) by hydrogen bonds from the guanidinium side chain.

The previous results have shown that the $O^* \rightarrow bR$ PT process is accompanied by the R82 upswing movement and the PRG formation. However, the results do not indicate at which stage of the PT process (at which ζ value) the two side processes become favorable and hence take place. In order to shed light on this issue, the end conformations of the 36 windows were analyzed with respect to their PRG and R82 conformation. For the R82 swing conformation this was accomplished by evaluating the distance of A44 to the R82 guanidinium group (like before in chapter 6, p. 47). With respect to Figure 6.7, the R82 guanidinium group can be seen in its swung-up orientation when the A44N-R82C ζ distance is smaller than 2.55 nm. For the detection of the PRG group formation the distance between the C δ atoms of E194 and E204 was used. With this definition, the PRG was formed when the E194C δ -E204C δ distance lies around 0.4 nm. On the other hand, the PRG complex was seen as not formed when this value was significantly larger than 0.45 nm in a simulation. These two properties of the window end conformations are shown in Figure 8.3 in dependence on the PT process (visualized by ζ). Concerning the PRG structure, the analysis indicates that this complex is formed already in a window where the ζ restraint was centered around -0.05. Also for the subsequent 24 windows, where $\zeta > -0.05$, the formed PRG is the dominating conformation. More precisely, 16 windows out of the 25 windows (64%) in this region exhibit the formed PRG. With these results in mind, it can be assumed that the PRG formation is an “early” side process of the PT reaction which happens directly after the formation of the proton hole. Moreover, since the majority of the later windows exhibit this conformation, the PRG formation is probably a more strongly exergonic process. Concerning the R82 orientation, Figure 8.3 clearly shows a preference for the swung-down conformation over the whole ζ range. Only in two windows, where the PT process is well advanced ($\zeta > 0.6$), the swung-up R82 orientation could be observed. Moreover, even in this late PT region ($\zeta = 0.6 - 0.85$) the observation of the swung-up R82 orientation is rare (only 25% of the windows are in this conformation). Thus, it can be concluded that the R82 upswing is a rather “late” PT side process, which probably is not as exergonic as the PRG formation.

The previous analysis has shown that the $O^* \rightarrow bR$ transition represents a multidimensional problem in collective variable space which could only roughly be described by the PT reaction itself. As a consequence, the PMF in Figure 8.2 represents presumably not the real free energy profile along the PT path and is probably distorted by the varying, non-equilibrated, synchronous degrees of freedom (PRG formation, R82 swing). In order to receive an equilibrated PMF along the $O^* \rightarrow bR$ transition it is necessary to keep control and equilibrate these essential degrees of freedom. One can achieve this by employing a multidimensional free energy approach like 3D-Metadynamics with ζ , the PRG formation and the R82 swing movement as the respective dimensions.

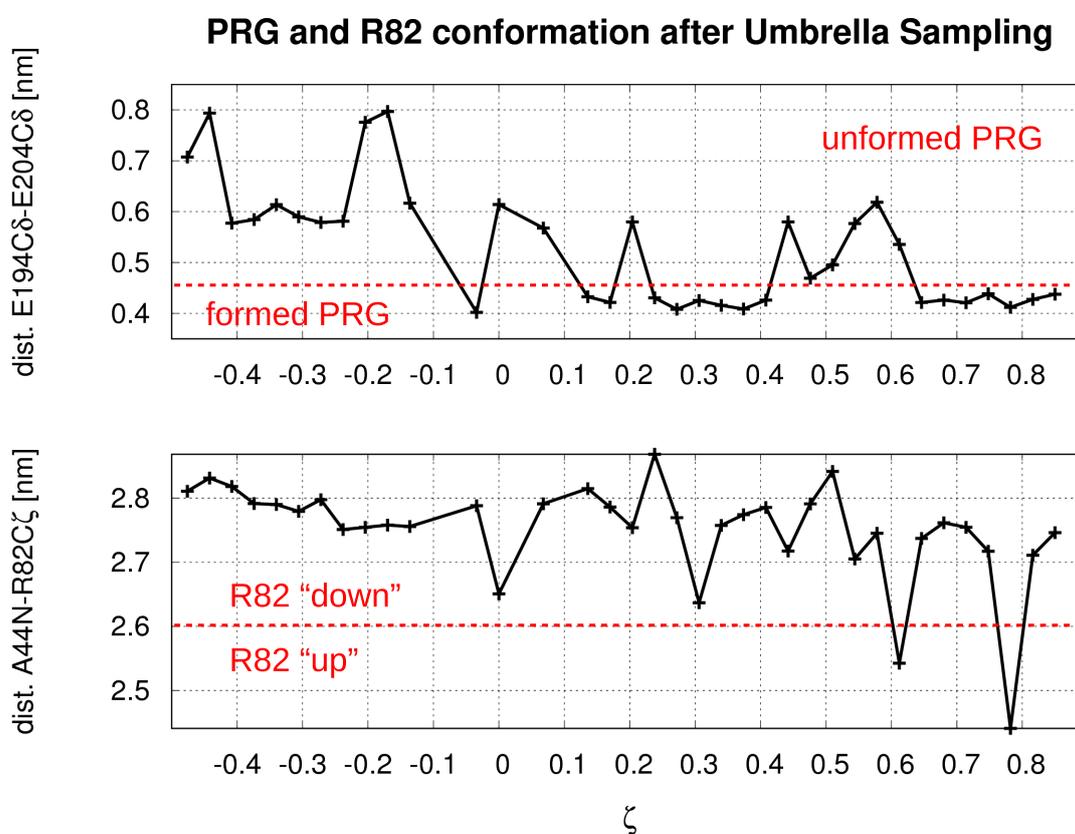


Figure 8.3: Analysis of the end conformations of the individual US windows. The R82 side chain orientation and the PRG conformation is depicted in dependence on the PT process (represented by ζ). Top: The PRG group is already formed in early stages during the PT process. After ca. $\zeta = -0.05$ the formation of the PRG occurs very often. Bottom: Contrary to the PRG formation, the R82 upswing movement is only observed in windows where the PT is well advanced. Even there, it is still only rarely observed.

8.3 O→bR: The Complete Reaction Mechanism

With the results from the O*→bR Umbrella Sampling calculations the thermodynamic gap between the O→bR transition can be closed. Together with the free energies from the O→O* (chapter 7, p.63) reaction, a complete free energy profile for the O→bR reaction can be constructed (see Figure 8.4). As previously shown, the O→O* reaction represents an endergonic reaction ($\Delta G_{O-O^*} = 4.2$ kcal/mol) with a reaction barrier of around $\Delta G_{O-O^*}^\ddagger = 8.8$ kcal/mol. In this reaction, the proton is transferred from D85 to D212 via a direct PT because the two amino acid side chains are close to each other. In the associated O*→bR transition, E204 is first protonated by a cavity water molecule such that a proton hole is formed, which then travels over a water wire towards the protonated D212 in order to get reprotonated by this amino acid. During this “inverse” PT the PRG is reformed and the R82 side chain reorients towards D85/D212. The transfer of the proton hole in combination with the structural rearrangements lead to a reaction barrier of $\Delta G_{O^*-bR}^\ddagger = 4.7$ kcal/mol. The recovery of the bR state represents, in contrast to the O→O* reaction an exergonic process, so that the bR state lies 1.2 kcal/mol lower on the Gibbs free energy profile than the O state. According to this, the O→bR transition represents a slightly exergonic reaction but still spontaneous process. The rate-limiting step on the free energy profile represents the O→O* reaction, which exhibits a nearly twice as high reaction barrier than the O*→bR reaction. Consequently, the O→bR transition time is mainly dominated by the D85 deprotonation step, whereas the proton hole transfer, PRG formation and R82 reorientation should happen much faster.

In order to estimate a rate for the O→bR transition we can employ transition state theory¹⁸⁶:

$$k = \frac{k_B T}{h} \left(1 - \exp \left[-\frac{h\nu}{k_B T} \right] \right) \cdot \exp \left[-\frac{\Delta G^\ddagger}{k_B T} \right] \quad (8.1)$$

where T is the temperature, ΔG^\ddagger is the free energy barrier, k_B is the Boltzmann constant, h is Planck’s constant and ν is the attempt frequency with which the TS occurs and can be furthermore approximated in our case by the vibrational frequency of the O-H bond:

$$\nu = \frac{1}{2\pi} \sqrt{\frac{k_{O-H}}{m_H}} \quad (8.2)$$

where m_H is the mass of the hydrogen atom and k_{O-H} is the force constant of the O-H bond. In the CHARMM22¹⁸⁷ force field, the force constant of the water O-H bond is $k_{O-H} = 450$ kcal/(mol·Å²). Consequently, one obtains a vibrational frequency of $\nu = 8.25 \cdot 10^{13}$ s⁻¹. Together with the $\Delta G_{O-O^*}^\ddagger$ of the rate-limiting step the reaction rate amounts to $k = 2.43 \cdot 10^6$ s⁻¹. This means in turn that the deprotonation of D85 happens on an average time scale of $\tau = 0.41$ μs and the O→bR transition would lie in the lower microsecond regime. However, in literature it is reported that the O→bR transition takes approximately 5 ms. With this, the calculated reaction barrier for the O→O* transition would be too low by roughly 5 kcal/mol. But one also has to keep in mind that the experimental estimate is quite rough, since the formation of the O state is difficult to identify distinctly.

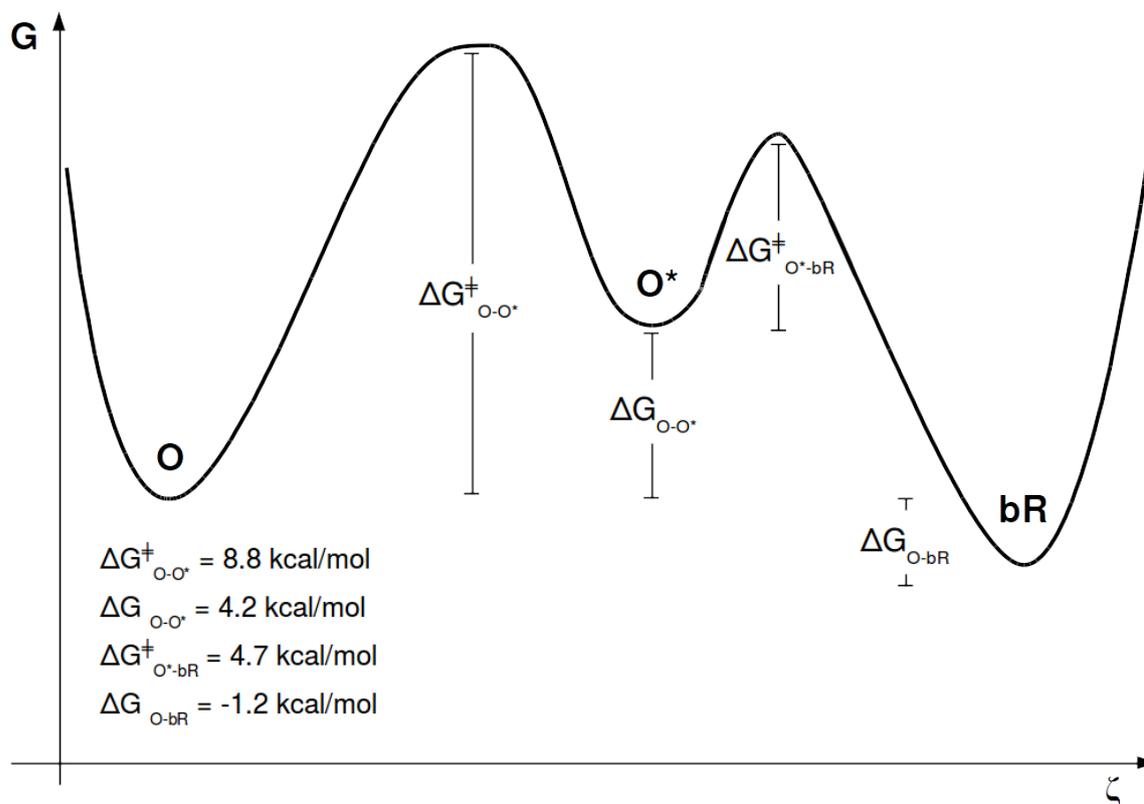


Figure 8.4: Proposed Gibbs free energy profile for the $O \rightarrow bR$ transition. The reaction can be divided into two steps: In the $O \rightarrow O^*$ transition, the proton is transferred from D85 to D212 (direct PT) in an endergonic reaction exhibiting a reaction barrier of 8.8 kcal/mol. In the subsequent $O^* \rightarrow bR$ step, a proton hole is formed close to the PRG region (while E204 gets protonated) and subsequently travels towards D212 in order to get reprotonated by this amino acid. This process exhibits a reaction barrier of 4.7 kcal/mol and makes the $O \rightarrow bR$ transition a slightly exergonic reaction in the end. The rate-limiting step of the whole process represents therewith the D85 deprotonation ($O \rightarrow O^*$).

8.4 Conclusions

This chapter aimed to resolve the thermodynamics and kinetics of the $O \rightarrow bR$ transition. Therefore, the assumed PT process between D212 and E204 (as part of the PRG) was investigated with the free energy method Umbrella Sampling. A first interesting finding was made with the observation where a proton was placed on a water between D212 and E204: During the equilibration, the hydronium exhibiting water wire was converted into a proton hole exhibiting water wire such that D212 and E204 got protonated. This finding was not unique. By contrast, it was repeatedly observed during the Umbrella Sampling runs. As a result, we can say with a good confidence that the PT process in $O^* \rightarrow bR$ shows a clear preference for a proton hole mechanism, which makes this PT an indirect one. The charge migration in the form of an OH^- ion is justified by the fact that the negative charge can easily pass by the positively charged R82 side chain and is furthermore stabilized by it.

These circumstances are not given when the charge travels as a hydronium ion. The proton hole mechanism is not an unexpected observation, because also experiments have shown that a mere mutation of D85 to serine converts bacteriorhodopsin into an anion (chloride) pump⁷². Consequently, the preference for pumping anions in this protein is given.

The PMF calculations indicated that the $O^* \rightarrow bR$ transition represents an exergonic reaction by $\Delta G_{O^* \rightarrow bR} = -5.4$ kcal/mol which is not only driven by the inverse PT process but also by the formation of the PRG and the R82 reorientation towards D85/D212. The PRG formation was detected in many individual Umbrella Sampling windows, where the PT process was rather close to the reactant state. Based on these observations, one can deduce that the PRG formation is a strongly exergonic reaction, which may ensure for the irreversibility of the transition. Secondly, the PRG formation can be seen as an “early” side process of the $O^* \rightarrow bR$ transition. In contrast to this, the R82 side chain orientation was only detected a few times in windows where the PT was well advanced. These facts lead to the assumption that the R82 upswing movement is probably only a slightly exergonic reaction and represents more a “late” side process in the $O^* \rightarrow bR$ transition where the proton hole is nearly protonated by the D212.

These two additional structural rearrangements complicate the $O^* \rightarrow bR$ transition in theory and furthermore maybe pose a technical problem to the WHAM analysis. The problem is that only one of the three essential degrees of freedom are controlled in the Umbrella Sampling simulations. As a consequence, there is a danger that the two other degrees of freedom were not sufficiently equilibrated and hence could affect the WHAM analysis and in turn the PMF profile.

Another side process in the $O \rightarrow bR$ transition, is the reformation of the HBN between K216RET, D85, D212 and the three water molecules. This process can only happen after the proton hole transfer is completed since then D212 is deprotonated and able to accept a hydrogen bond. However, this reformation was not equilibrated in the “late” Umbrella Sampling windows. With the thermodynamic data of the $O^* \rightarrow bR$ transition a complete free energy profile along the PT process was proposed. On this profile, the $O \rightarrow O^*$ transition and hence the deprotonation of D85 represents the rate-limiting step. This prediction is in accordance to pH jump experiments⁶⁴. Furthermore, this endergonic transition is directly coupled to the exergonic $O^* \rightarrow bR$ transition, which in turn makes the complete $O \rightarrow bR$ transition slightly exergonic by $\Delta G_{O \rightarrow bR} = -1.2$ kcal/mol. This prediction deviates from experimental reports, which state that the $O \rightarrow bR$ transition is strongly downhill due to the regeneration of the low D85 pK_a (ca. 2.5) and the high PRG pK_a (ca. 9)²³. It is conceivable that an appropriate treatment of the poorly converging degrees of freedom (especially the HBN formation and the R82 swing) will lead to a more exergonic free energy profile.

Also, the estimate of the $O \rightarrow bR$ transition time of $\tau = 0.41$ μ s is considerably faster than the literature value of ca. $\tau = 5$ ms. However, the difficult experimental conditions of the $O \rightarrow bR$ transition also leave room for discussion of this literature value.

Parametrization and Benchmarking of Delta-Pauli DFTB

The performance of DFTB-D3 with respect to reaction barriers was already touched in chapter 7, p.74 where the semiempirical method was benchmarked in comparison with the hybrid-DFT method B3LYP. The results indicated that DFTB-D3 describes the PT qualitatively correct, but the method suffers from a small underestimation compared to the DFT reference with respect to reaction barriers. It is known that this deviation results to a certain extent from the underestimated short-range Pauli repulsion in DFTB, resulting mainly from the minimal basis set approximation and from inherited deficiencies of PBE. The barrier of intermolecular reactions is only one observable among others which is affected by the inappropriate description of this non-covalent interaction. Also intramolecular barriers around a torsional angle are underestimated by a few kcal/mol. Consequently, this error can also be found in conformational transitions of biomolecular systems like peptides (e.g. ϕ/ψ angles) etc. Moreover, the missing Pauli repulsion also affects structural properties of molecules like equilibrium distances and spatial orientation. Again, this error is also found in properties of large molecular complexes like for example in a slightly deviating radial distribution function for water.

In order to correct for this missing short-range repulsion, Maximilian Kubillus developed an empirical correction model named Delta-Pauli and implemented it in the DFTB+¹⁸⁸ software package. Until now, no parameter set is available for the Delta-Pauli model. Thus, in this chapter a first parameter set will be derived and subsequently benchmarked.

9.1 A Fitting Approach for Delta-Pauli DFTB

It has been shown in chapter 2, p.11 that a non-covalent dissociation curve consists of two significant parts: the short-range Pauli repulsion area and the attractive dispersion-controlled region. Both interactions smoothly pass into each other. In order to achieve a good balance between the D3 dispersion and Delta-Pauli correction, it is good practice when deriving parameters for Delta-Pauli also to refit the D3 dispersion parameters. The Delta-Pauli energy (see equation 3.30, p. 11) contains atom-specific (s_X) as well as universal

parameters (ξ , ζ , α , d), whereas the empirical D3BJ dispersion consists of three parameters: a_1 , a_2 and s_8 . For biological systems, which mainly comprise C, H, N and O, this leads to eleven parameters in total which need to be fitted.

In this approach, the Delta-Pauli parameters were derived with the help of a training set-based fitting protocol. Thus, it was necessary to design an appropriate training set with systems that exhibit the mentioned types of non-covalent interactions. As fitting method PSO¹²⁶ was used. The goal of the fitting procedure was to derive a DP/D3 parameter set that yields a more repulsive behavior than DFTB3/3OB-D3BJ for short-range non-covalent interactions. At the same time these parameters should maintain the good performance of the D3 dispersion correction.

9.1.1 Training Set Design and PSO Fitting

The first part of the training set consisted of the S66x8 set designed by Řezáč and coworkers⁹⁸. This benchmark set was originally designed to cover most of the non-covalent interaction motifs occurring in CHNO-biomolecular systems. Moreover, the authors tried to keep a good balance between electrostatic and dispersive non-covalent interactions. All in all the S66x8 set consists of 66 biomolecular dimer-complexes and each was calculated at the equilibrium distance as well as at 7 non-equilibrium distances to recover the dissociation curve. Previously, the 594 configurations were optimized at the MP2/cc-pVTZ level and their energies were calculated at the CCSD(T)/CBS level. All configurations and the corresponding energies were taken from the BEGDB¹⁸⁹ website. Since hydrogen-bonded systems pose a problem to Delta-Pauli (cf. subsection 9.1.3), OH- and NH-hydrogen-bonded systems were excluded from the training set.

In order to specifically train the Delta-Pauli model to correct for the underestimation of rotational barriers, we created the R10 set. It is based on all possible non-redundant CHNO-combinations of ethane (see Appendix, Figure B.1), leading to ten small molecules in total. For each system, non-redundant rotational conformers in 10° steps were created with the Gaussian09¹⁹⁰ quantum chemistry package. A restrained optimization at the B3LYP-def2TZVP(D3-BJ) level followed, where the torsional angle was kept constant. From these minimized rotamers SCS-MP2 energies were obtained. For the two latter steps the Turbomole v.7.0¹⁹¹ quantum chemistry package was used. The modified S66x8 and R10 set led to 476 data points (see Table 9.1).

Table 9.1: Detailed composition of the training set that was used to derive parameters for Delta-Pauli and D3.

Training Set	Description	Data points	Level of theory
S66x8 ⁹⁸	Dimer interaction energies along a dissociation curve.	344	CCSD(T)/CBS
R10	Relative energies of non-redundant, ethane-derived rotamers.	132	SCS-MP2

The Delta-Pauli and D3 parameters ξ , ζ , α , d , s_C , s_H , s_N , s_O , a_1 , a_2 , s_8 were derived via PSO fitting against the R10 and S66x8 set in a predefined parameter space (see Table B.2). A PSO implementation was used which was written in Python by Maximilian Kubillus. For the algorithm the settings reported in Table B.1 were used. The fit procedure started with an initial global best fitness of 1.26. After 90 iterations this fitness value converged to 1.07 and yielded the parameters shown in Table 9.2. It is noteworthy that the s_H and a_1 parameter are at their fit boundaries.

Table 9.2: Obtained Delta-Pauli and D3 parameters, after fitting to the training set depicted in Table 9.1.

Parameter	Value
ξ	29.33
d	3.21
ζ	2.15
α	0.75
s_C	9.49
s_H	0.11
s_N	46.06
s_O	44.51
a_1	0.10
a_2	8.48
s_8	9.15

9.1.2 First Tests with the Delta-Pauli Parameters

In order to determine the positive and negative effects of Delta-Pauli, the R10 and S66x8 structures were reevaluated via single point energy calculations with the above listed parameters (see Table 9.2) on basis of the CCSD(T)/SCS-MP2 geometries.

Figure 9.1 depicts two rotamer systems from the R10 set and two dimer dissociation systems from the S66x8 set. For the rotamer systems the single point energies are plotted against the dihedral angle. For the dissociation systems the single point energies were plotted against the ratio of the actual dimer displacement r with respect to the equilibrium displacement r_0 . The four examples in Figure 9.1 represent positive and negative cases of Delta-Pauli respectively: The top, left panel displays the rotational barrier of ethanediol. Obviously, DFTB3-D3 underestimates this barrier by several kcal/mol. For instance, the 0° conformation, where the two terminal O atoms stand eclipsed to each other yields approximately 4 kcal/mol deviation with respect to the SCS-MP2 reference. Also the second maximum barrier height at 120° where each oxygen is in eclipsed conformation towards hydrogen is underestimated by around 3 kcal/mol with DFTB3-D3. Contrary to this, the rotational barrier energies improve with Delta-Pauli DFTB (DP-DFTB): The deviation of the first maximum from the reference energy is only ca. 1 kcal/mol lower than the reference energy. The same applies to the second maximum, where the rotational barrier is overes-

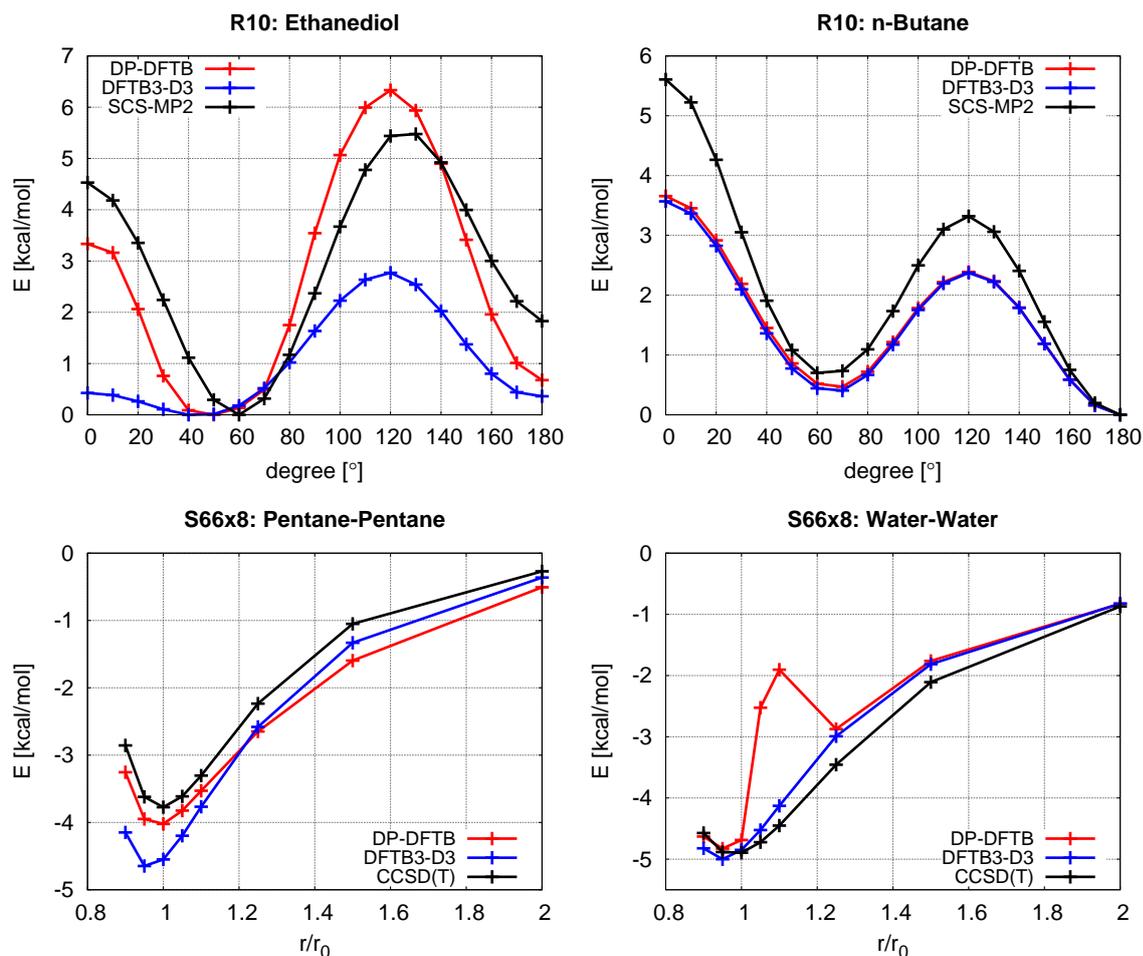


Figure 9.1: Top: Rotational barriers of ethanediol (left) and n-butane (right) calculated with SCS-MP2, DFTB-D3 and DP-DFTB. Bottom: Dissociation curves of the pentane (left) and water (right) dimer calculated with CCSD(T), DFTB-D3 and DP-DFTB. For DP-DFTB the original parameters listed in Table 9.2 were used.

estimated by Delta-Pauli by around 1 kcal/mol. However, for n-butane (top, right panel in Figure 9.1) we don't observe any improvement of the rotational barrier with Delta-Pauli: The results are identical to the DFTB3-D3 energies, where the energies are underestimated by approximately 1 – 2 kcal/mol for the 0° and 120° conformation.

The bottom, left graph in Figure 9.1 shows the dissociation curve for the pentane dimer of the S66x8 set. The results verify that the Delta-Pauli method with its modified D3 parameters is able to maintain the performance of the standard D3 method and yields a physically correct image of the dissociation curve. Moreover, the pentane-pentane dimer is too stable in the DFTB-D3 representation, which is not the case for DP-DFTB. When we have a closer look at the minimum of the interaction energy, we can identify a slight difference between the DFTB-D3 and Delta-Pauli results: The DFTB-D3 data points at $r/r_0 = 0.90$ and 1.00 are nearly isoenergetic. This is not the case with Delta-Pauli: The

data point $r/r_0 = 0.90$ is slightly higher than for the one with $r/r_0 = 1.00$ so that the dissociation curve shape is more similar to the one derived by CCSD(T). This effect stems from the improved repulsive behavior introduced by the Delta-Pauli model. The dissociation curve for the water dimer system in the bottom, right graph depicts an example, where the Delta-Pauli model fails completely. The graph exhibits a strong peak from $r/r_0 = 1.0$ to 1.25 leading to a destabilization of the complex in this area. Such a behavior is totally wrong and diverging from the reference and also from the DFTB3-D3 results. Moreover, this type of artifact is also observed in all other systems that comprise hydrogen bonds (e.g. AcOH dimer, MeNH₂-MeOH complex etc.). The same problem applies to weakly hydrogen-bonded complexes (e.g. benzene-AcNH₂(NH- π), ethyne-AcOH(OH- π) etc.).

9.1.3 Hydrogen Bonds – A Pitfall for Delta-Pauli

Hydrogen bonds are probably the most important non-covalent interactions and their physically correct description is crucial for the exact prediction of biomolecular systems of interest like peptides, sugar conformers etc. Thus, it is indispensable to get rid of those repulsive peaks which Delta-Pauli introduces to hydrogen-bonded systems (see Figure 9.1).

Let us have a closer look on the water dimer system depicted in Figure 9.2A, left. The picture shows two overlaid conformations of the S66x8 water dimer at two different points of the dissociation curve: $r/r_0 = 1.00$ and $r/r_0 = 1.25$. The peak in the dissociation graph in Figure 9.1 arises exactly between these two conformers, which exhibit hydrogen bonds (the O-H distance lies between 3.8 – 4.7 a.u.). A closer look on the functional behavior of the switching function for the O-H atom pair (Figure 9.2A, right) reveals that Delta-Pauli switches on/off in this hydrogen-bonded region. The effect is the artificial disruption of the hydrogen bond which we observe as a peak in the dissociation curve. This behavior is expected, since the Delta-Pauli model is not able to “detect” hydrogen bonds. It always gets activated/deactivated at the same atom pair distance, no matter if a hydrogen bond is present or not. The same circumstance applies to N-H hydrogen-bonded complexes like peptide-MeNH₂, water-pyridine etc.

Moreover, systems which form so-called weak hydrogen bonds also suffer from this issue. A weak hydrogen bond is qualitatively similar to a normal one but it is based on the interaction of an electron withdrawing group X-H (X=O,N,C, ...) with the electrons of a delocalized π -system. Thereby, the interaction strength grows with the acidity of the X-H group¹⁹². Figure 9.2B, left shows the benzene-water(π -OH) complex in its equilibrium state ($r/r_0=1.00$). The distance between the water hydrogen atom that is facing the benzene ring and the center of mass of the benzene carbon atoms is 4.77 a.u. This distance is smaller than ordinary C-H non-covalent interactions due to the stabilizing interaction of the O-H group with the benzene π -system. According to the C-H switching function in Figure 9.2B, right, Delta-Pauli would also disrupt these weak C-H hydrogen bonds.

As a consequence, to make the actual Delta-Pauli implementation work for hydrogen-bonded systems, it is necessary to exclude the O-H, N-H and C-H atom pairs from this correction. In order to achieve this, the Delta-Pauli cutoff radii r_{ab}^0 for these atom pairs were manually adjusted such that the Delta-Pauli correction switches on/off at distances

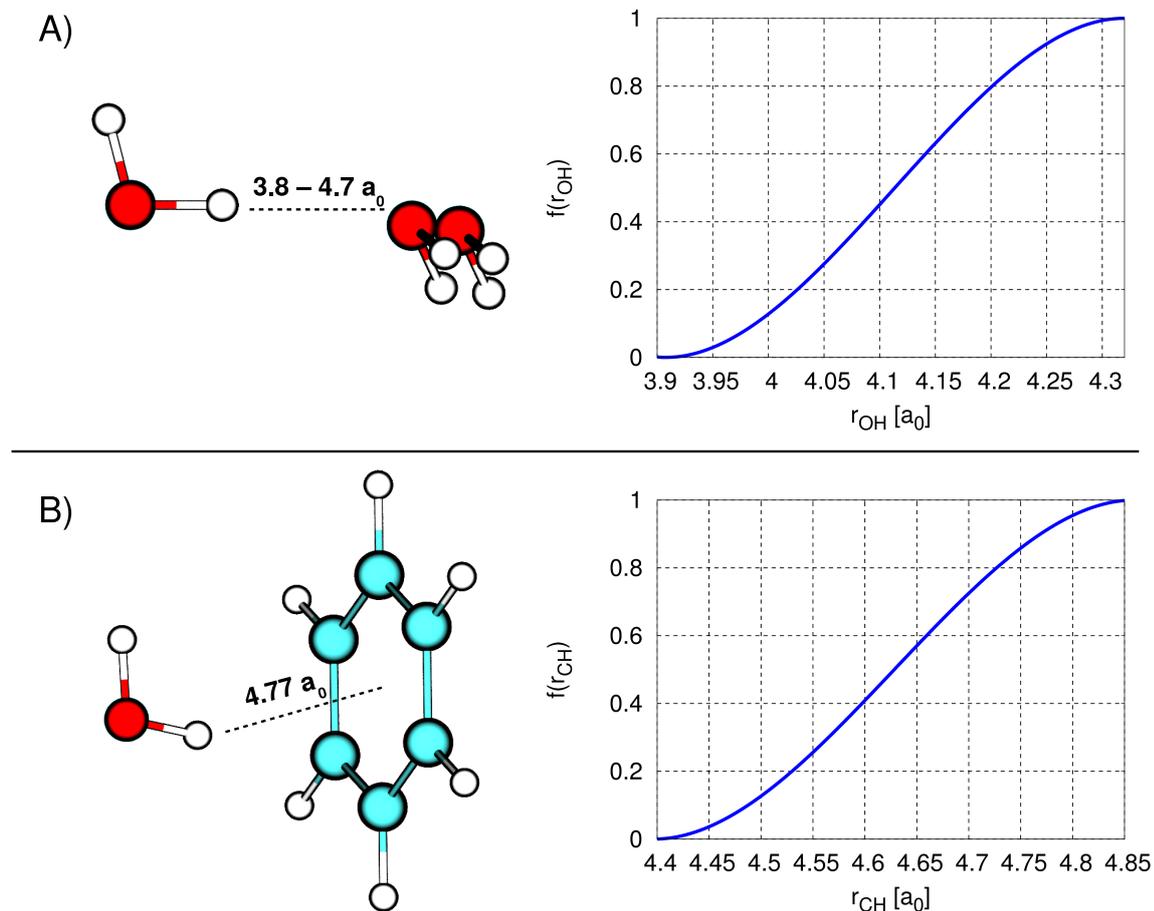


Figure 9.2: A, left: Structure of the water dimer for the data points $r/r_0 = 1.00$ and 1.25 . The O-H distance between the water molecules is indicated. Right: Functional behavior of the O-H switching function. B, left: Structure of the benzene-water(π -OH) complex in the equilibrium geometry $r/r_0 = 1.00$. The distance between the water hydrogen to the center of mass of the benzene ring is indicated. Right: Functional behavior of the C-H switching function.

where hydrogen bonding is not present anymore. Hence, these cutoff radii were extended by $1.0 a_0$. Since at such large distances the $V_{\text{rep}}(r_{ab})$ approaches anyway zero, the Delta-Pauli repulsion is essentially deactivated for the O-H, N-H and C-H atom pairs.

Nevertheless, the manual adjustment of the X-H cutoff radii seems to be the only workaround for this issue, which does not need any modification of the underlying theory (introduction of hydrogen bond recognition etc.).

Accordingly, from now on, we will use the modified Delta-Pauli cutoff radii r_{ab}^0 listed in Table 9.3.

9.1.4 Manual Correction of the C/H Overlap Proportionality Parameters

Another issue, which should be addressed, is the poor performance of Delta-Pauli for the n-butane rotational barrier shown in Figure 9.1. Fit parameters which were derived on a training set that consisted of only CH systems reproduced this barrier much better (not shown here). Thus, the parameters depicted in Table 9.2 yield a too low repulsion for the C-H interaction. This fact is underpinned when we have a look at the fitted overlap proportionality parameters. These parameters should be proportional to the atom size, but the fitted s_C and s_H parameters are remarkably smaller than the ones for oxygen and nitrogen. Moreover, the s_H value is so small that it is close to the lower fit boundary of 0.1. Test calculations on the training set, where the s_C and s_H parameters were slightly increased manually yielded a better overall performance. Therefore several s_C/s_H combinations were evaluated to gain more repulsion for rotational barriers without compromising dissociation curves for CH systems. Eventually, the s_C and s_H parameters were manually increased to the parameters shown in Table 9.3.

Table 9.3: List of Delta-Pauli cutoff radii for the atom pairs O-H, N-H, C-H and overlap proportionality parameters for C and H. The original and modified parameters are shown respectively.

Parameter	Original value	Modified value
r_{OH}^0	4.113	5.113
r_{NH}^0	4.472	5.472
r_{CH}^0	4.628	5.628
s_C	9.49	10.50
s_H	0.11	1.50

9.2 Performance of Delta-Pauli on Training Set Systems

9.2.1 Single Point Energies

With the modified $r_{OH}^0, r_{NH}^0, r_{CH}^0, s_C$ and s_H parameters, the training set data were recalculated in order to evaluate what effects the new parameters yielded for the rotational barriers and dissociation curves. Figure 9.3 shows the same systems like before (see Figure 9.1) but calculated with the updated Delta-Pauli parameters from Table 9.3. The most obvious change is the disappearance of the peak in the water dimer dissociation graph. The second finding is, as desired, a more repulsive behavior for the CH systems, so that the rotational barrier of n-butane is slightly improved. But it is also noticeable that the repulsive parameters s_C and s_H have to be carefully adjusted such that sensitive, short-range data points in the dissociation graphs (e.g. $r/r_0 = 0.9$ in the Pentane dimer) do not get too repulsive. Since the mean pitfalls seem to be fixed, we can go on with the analysis of the whole training set. For the sake of clarity, all rotational barriers and dissociation graphs have been attached to the Appendix (see Figure B.2, p.132). Figure B.3 shows the mean absolute deviation (MAD) of DFTB3-D3 and DP-DFTB for each system in the R10/S66x8 training set. The results indicate that Delta-Pauli decreases the error of the rotational barriers compared to DFTB-D3 without exception. This fact is confirmed by the averaged MADs shown in Table 9.4 where DP-DFTB exhibits a 0.4 kcal/mol lower MAD for rotational barriers than DFTB-D3. Furthermore, with respect to Figure B.3 one can say that Delta-Pauli has a very positive effect on rotational barrier systems that depend on H-O, O-O, H-N, N-N and N-O contacts (e.g. ethanol, ethanolamine etc.). However, for systems where the rotational barrier stems mostly from H-H, C-H, C-C, C-N and C-O contacts (e.g. n-propanol, n-butane etc.) the improvements are not so huge.

Considering the S66x8 systems, it can be observed that DP-DFTB performs a little worse than DFTB-D3 so that the overall MAD from the reference is for DP-DFTB by 0.1 kcal/mol higher than for DFTB-D3 (see Figure 9.4). With respect to Figure B.3, there is no obvious system-dependent pattern recognizable but one can say that the biggest outliers of DP-DFTB stem from pure π - π -bonded CH systems like the benzene(π - π) dimer. Despite the higher MAD of DP-DFTB, a closer look on the dissociation plots reveals significant improvements of DP-DFTB compared to DFTB3-D3: Some systems like the pentane-AcNH₂ or the neopentane dimer have no minimum in the DFTB3-D3 representation whereas DP-DFTB solves this problem. Furthermore, for some systems (e.g. pyridine-uracil(π - π), ethene-pentane etc.) the minimum of the dissociation curve is shifted with DFTB3-D3 to smaller distances compared to the reference. This overbinding effect is mostly corrected in the DP-DFTB representation.

To sum up, it has been shown that DP-DFTB outperformed DFTB-D3 regarding the description of rotational barriers. Furthermore, the reasonable performance of the D3 dispersion correction has been conserved in Delta-Pauli, and the overbinding has been compensated partially in some cases.

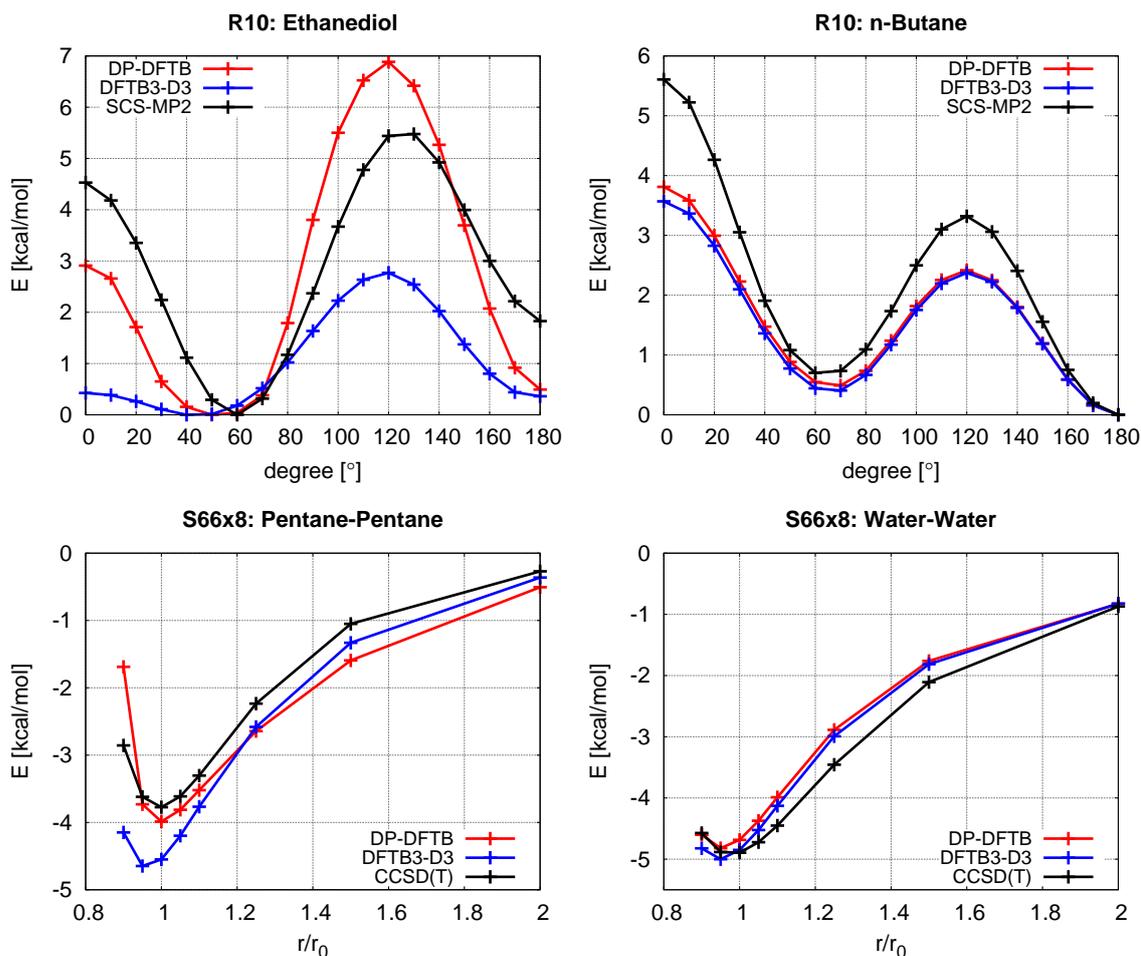


Figure 9.3: Top: Rotational barriers of ethanediol (left) and n-butane (right) calculated with SCS-MP2, DFTB-D3 and DP-DFTB. Bottom: Dissociation curves of the pentane (left) and water (right) dimer calculated with CCSD(T), DFTB-D3 and DP-DFTB. For DP-DFTB the modified parameters listed in Table 9.3 were used.

9.3 Analysis of the Delta-Pauli Repulsion

In the previous section it has been shown that DP-DFTB sufficiently introduces repulsion to short-range interatomic distances. Nevertheless, we do not know

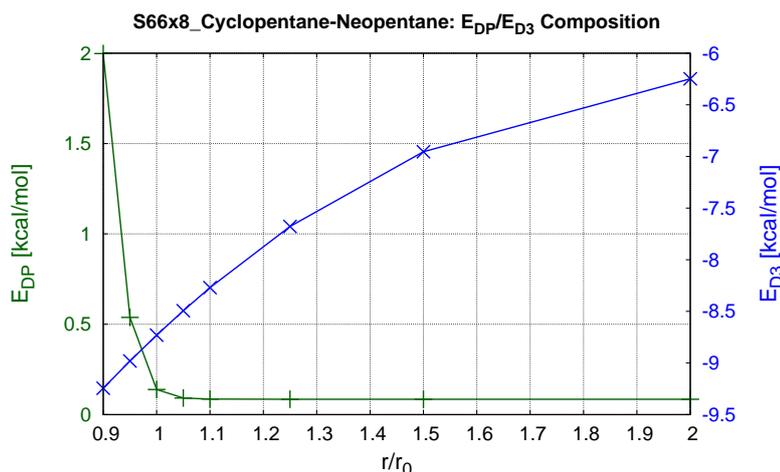
- how the D3 dispersion energy and the Delta-Pauli repulsive energy behave in detail,
- which atom species introduce repulsion and to what extent,
- if Delta-Pauli is geometry-sensitive and
- how geometries change with Delta-Pauli.

These are the questions we want to address in this section.

Table 9.4: Averaged MADs of DFTB-D3 and DP-DFTB for the R10 and S66x8 training set.

Set	DFTB-D3 [kcal/mol]	DP-DFTB [kcal/mol]
R10	0.86	0.40
S66x8	0.73	0.81

Figure 9.4 shows the D3 and Delta-Pauli energy contribution for the cyclopentane-neopentane complex over a dissociation range from $r/r_0 = 0.9 - 2.0$. The D3 energy decreases continuously from $r/r_0 = 2.0 - 0.9$, which means that over the whole range the D3 energy acts stabilizing to the molecule complex. Contrary to this, the Delta-Pauli energy decays exponentially in a narrow interval from $r/r_0 = 0.9 - 1.1$. It is noteworthy that the Delta-Pauli energy decays not fully to zero in the long-range part, but rather applies a constant repulsive offset of 0.08 kcal/mol to the system. Compared to D3 with $-6.2 - -9.2$ kcal/mol, the Delta-Pauli energy lies with $0.08 - 2.0$ kcal/mol on a much smaller scale. These results prove that Delta-Pauli only affects the short-range repulsive part, like we wanted. The medium- and long-range part of the non-covalent interactions are still governed by the D3 dispersion correction. With the help of Figure 9.5 we can address

**Figure 9.4:** Contribution of the Delta-Pauli (E_{DP}) and D3 dispersion energy (E_{D3}) for the cyclopentane-neopentane complex.

the questions which atom species introduce repulsion and if the Delta-Pauli correction is geometry-sensitive. Depicted are the three largest Delta-Pauli energy contributions for the ethanolamine system in dependence on the O5-C1-C2-N9 dihedral angle. The biggest contribution to the Delta-Pauli energy comes from the 1,4-interaction of O5 and N9 in the 0° conformation since there the two atoms are directly facing each other. The repulsive interaction of this atom pair switches off with increasing dihedral angle, so that the Delta-Pauli energy contribution for this atom pair vanishes at 70° since the distance $r_{O_5N_9}$ becomes too large. The second biggest contribution is due to the 1,3-interaction of the O5 with the

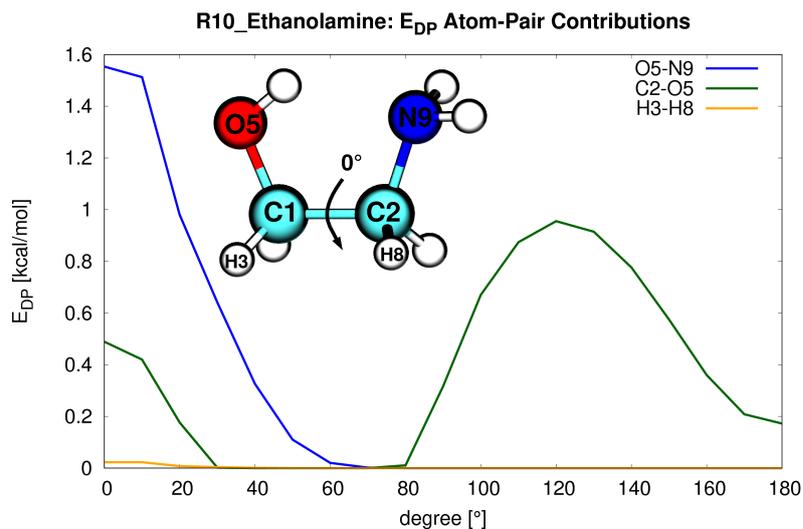


Figure 9.5: Contribution to E_{DP} of the three “most repulsive” atom pairs in ethanolamine in dependence on the dihedral angle O5-C1-C2-N9.

C2 atom. The repulsive energy for this atom pair exhibits two maxima at 0 and 120°. Between these two significant data points the Delta-Pauli energy vanishes completely in the region between 30 – 70°. Since the distance r_{O5C2} is nearly constant during rotation, this fluctuation in E_{DP} must be caused by the shielding effect of neighbouring hydrogen atoms. An interesting fact is that the third biggest contribution to the Delta-Pauli energy stems not from the atom pair N9-C1 as one would assume since this interaction is comparable to the C2-O5 interaction. Instead, the third biggest contribution results from the 1,4-interaction of the H3 atom with the H8 atom. This interaction vanishes after its maximum at 0° very quickly. Furthermore, compared to the two other contributions this interaction is much weaker. Up to now, only energies have been considered but we do not know if Delta-Pauli is also able to correct for geometries. Therefore, since analytical gradients are still under development, numerical gradients for Delta-Pauli were used to optimize the S66x8 geometries. With the minimized structures the root-mean-square deviation (RMSD) from the CCSD(T) reference structures were calculated and the results were again compared to DFTB-D3. Figure 9.6A depicts the structural RMSD values of DP-DFTB and DFTB-D3. Like for energies, the deviation of DP-DFTB from the reference is smaller on average than with DFTB-D3 but not for every individual system. Moreover, for systems where DP-DFTB improved energies (e.g. pentane dimer, pyridine-ethene complex etc.) also an improvement in the overall geometry is the consequence. This fact is confirmed by Figure 9.6B. It shows the superimposed reference-, DFTB-D3- and DP-DFTB-optimized structures of the pentane dimer and the pyridine-ethene complex respectively. The geometries indicate that the DP-DFTB proposed structure is nearly overlapping with the reference, whereas the structure proposed by DFTB-D3 shows a larger deviation.

To conclude, it has been shown that Delta-Paulis screening function yields reasonable results which makes the correction geometry-sensitive. The atom pair specific analysis of

the Delta-Pauli energy proves that the biggest part of the repulsive energy originates from the heavy atoms C, N and O. Furthermore, in this section it has been shown that for systems where energies improved with DP-DFTB, also the corresponding geometries got better.

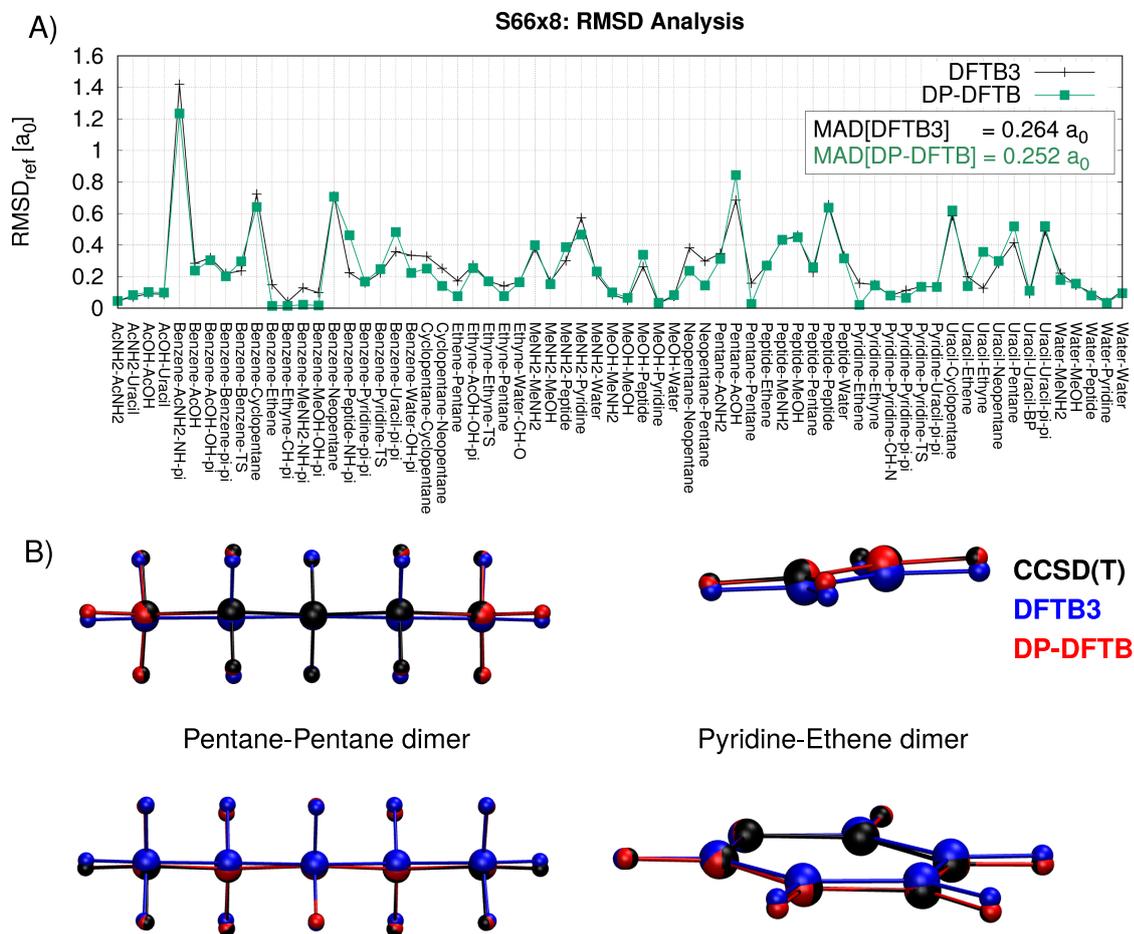


Figure 9.6: A: Structural RMSD analysis for the S66x8 systems. The RMSD value is defined by the deviation of the DFTB-D3- and DP-DFTB-optimized geometries from the CCSD(T) reference in units of a_0 . The average MAD of the RMSD is smaller for DP-DFTB compared to DFTB-D3. B: DFTB-D3- and DP-DFTB-optimized geometries were superimposed with the CCSD(T) reference for the pentane dimer and pyridine-ethene dimer. The deviation from the reference is bigger for DFTB-D3 as for DP-DFTB.

9.4 Performance of Delta-Pauli on Benchmark Systems

In the previous section it has been shown that the Delta-Pauli extension successfully improved the short-range repulsive behavior of DFTB-D3. However, there still remains the open question, if this improvement only applies to training set systems or if this correction also improves the description of molecules that were not part of the training set. In order to

check for this, we collected systems for our benchmark set that are of organochemical and biological relevance. Furthermore, a prerequisite for the benchmarking of DFTB3-D3 and DP-DFTB is that those systems should have been calculated with a reliable higher-level theory (e.g. coupled cluster, composite methods) or should be backed up by experimental results. One proper source is the non-covalent branch of Grimme’s GMTKN30 set¹⁷⁴. It consists of benchmark sets like the ACONF, ADIM6, IDISP, SCONF, WATER27 and PCONF subset. These include relative energies of alkane chains, alkane dimers, dispersion dominated systems, sugar conformers, tripeptide conformers and water clusters. It has to be mentioned that for the WATER27 data the revised energies by Manna et al.¹⁹³ were taken. Furthermore, we used the extended PCONF(FGG) subset¹⁹⁴ that was recently recalculated¹⁹⁵. Another appropriate set is the L7 set¹⁹⁶, which contains interaction energies of large complexes that are mostly dispersion-stabilized. A further objective was the benchmarking of DP-DFTB on two model systems which are commonly used to assess the performance of quantum chemical methods for proteins and nucleic acids respectively: the alanine dipeptide and sugar pucker conformers. For the alanine dipeptide, geometry-optimized data from our group⁹² was used where the ϕ/ψ angles were scanned in 6° steps in order to recover the PES. For the sugar pucker conformers (dA, dG, dC, dT, rA, rG, rC and rU) the PES of the two endocyclic torsions Z_x and Z_y was also scanned in 6° steps. The geometry-optimized data was obtained from Huang et al.¹⁹⁷. The complete benchmark set with further description is listed in Table 9.5.

Table 9.5: Composition of the benchmark set for DP-DFTB.

Benchmark Set	Description	Level of theory
ACONF ¹⁷⁴	Relative energies of C_nH_{2n+2} alkane conformers.	W1h-val
ADIM6 ¹⁷⁴	Dimer interaction energies of n-alkanes.	CCSD(T)/CBS
IDISP ¹⁷⁴	Intramolecular dispersion interactions of large organic systems.	Theo. + exp.
L7 ¹⁹⁶	Large (48-112 atoms) dispersion-stabilized complexes.	CCSD(T)/CBS
SCONF ¹⁷⁴	Relative energies of D-galactitol and D-glucopyranose sugar conformers.	CCSD(T)/CBS
FGG(PCONF) ^{194,195}	Relative energies of Phe-Gly-Gly tripeptides.	CCSD(T)/CBS
WATER27 ¹⁹³	Relative energies of $(H_2O)_n$, $H^+(H_2O)_n$ and $OH^-(H_2O)_n$ clusters.	CCSD(T)/CBS + MP2/CBS
Ala-Dipeptide ⁹²	Scan of the ϕ/ψ PES.	BLYP/def2-QZVP-D3BJ
Sugar pucker ¹⁹⁷	Scan of the Z_x/Z_y PES.	MP2/6-311++G(3df,2p) //MP2/6-31++G(d,p)

9.4.1 Single Point Energies

The energies of ACONF, ADIM6, IDISP, L7, SCONF, FGG(PCONF) and WATER27 were recalculated with DFTB-D3 and DP-DFTB. The y-axis of the corresponding bar charts displays the deviation of these methods from the reference. For comparison, precalculated results of two other DFT-functionals (preferably GGA) were listed.

ACONF

The bar chart for the ACONF subset in Figure 9.7 shows clearly that the reaction energies are better described by DP-DFTB compared to DFTB-D3. More precisely, the mean absolute deviation of DFTB-D3 for this subset is 0.71 kcal/mol and gets lowered by half with the Delta-Pauli correction to 0.34 kcal/mol. Furthermore, it can be observed that compared to the CCSD(T) reference the two tight-binding models suffer in each reaction from a barrier-underestimation. Contrary to this, the GGA density-functional methods BLYP and PBE overestimate the barrier for nearly every reaction. Furthermore, the DFT methods outperform DFTB-D3 in every ACONF reaction. This is not the case for DP-DFTB (e.g. PTT→PGG, Httt→Hggg, Httt→Htgg). A closer look on the composition of the ACONF reactions reveals why Delta-Pauli performs better compared to DFTB-D3: This subset consists of torsional rearrangement reactions of alkane conformers¹⁹⁸, for example trans-n-butane → gauche-n-butane. The barrier-underestimation of DFTB-D3 results from the missing short-range repulsion. DP-DFTB was especially fitted to reproduce such reactions (see R10 training set) by introduction of the missing repulsion.

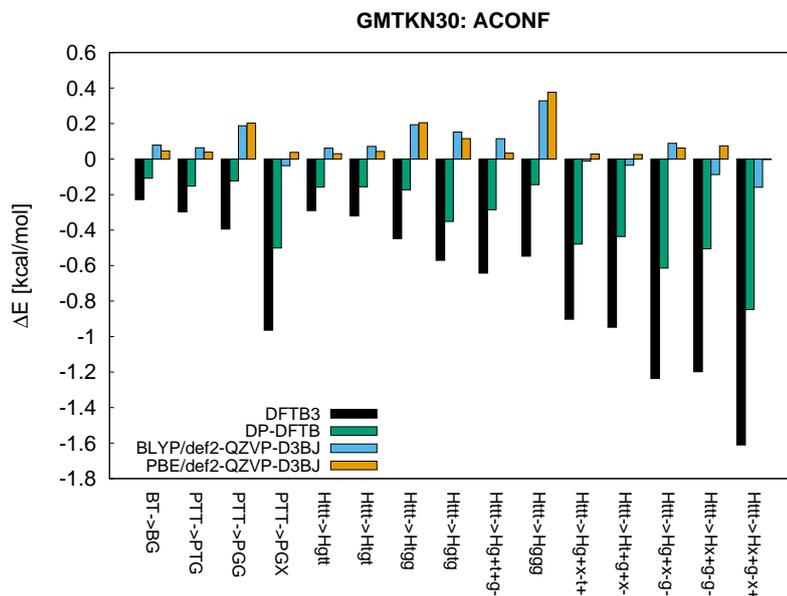


Figure 9.7: Deviation of DFTB-D3 and DP-DFTB from the W1h-val reference energies for the ACONF reactions. The overall MAD for DFTB-D3 amounts to 0.71 kcal/mol and is decreased with DP-DFTB to 0.34 kcal/mol. For comparison also the deviation of BLYP/def2-QZVP-D3BJ and PBE/def2-QZVP-D3BJ are listed.

ADIM6

The positive effects of Delta-Pauli also hold true for the ADIM6 results shown in Figure 9.8. The MAD from the reference is decreased from 0.75 kcal/mol (DFTB-D3) to 0.24 kcal/mol with Delta-Pauli. With one exception, PBE-DFT performs always better than DFTB-D3. On the other hand, DP-DFTB outperforms its mother-functional in every case. The BLYP functional exhibits, compared to the other methods, the biggest deviation from the reference. Furthermore, like with DFTB-D3, the error grows with increasing system size. Let us connect the mentioned numbers with the composition of this benchmark set: The ADIM6 set consists of intermolecular, London dispersion interaction energies for alkane dimers. These interactions are not so small as one would assume. Consequently, the interaction energy of the n-hexane/n-heptane is on the same magnitude as the hydrogen bond of the water dimer¹⁹⁹. For a proper description of these interactions, a dispersion correction like D3BJ is indispensable. The results show that the parametrized D3 parameters of DP-DFTB yield sensible results for this type of non-covalent interaction. However, one has to mention that the corrected energies lie on a very small scale.

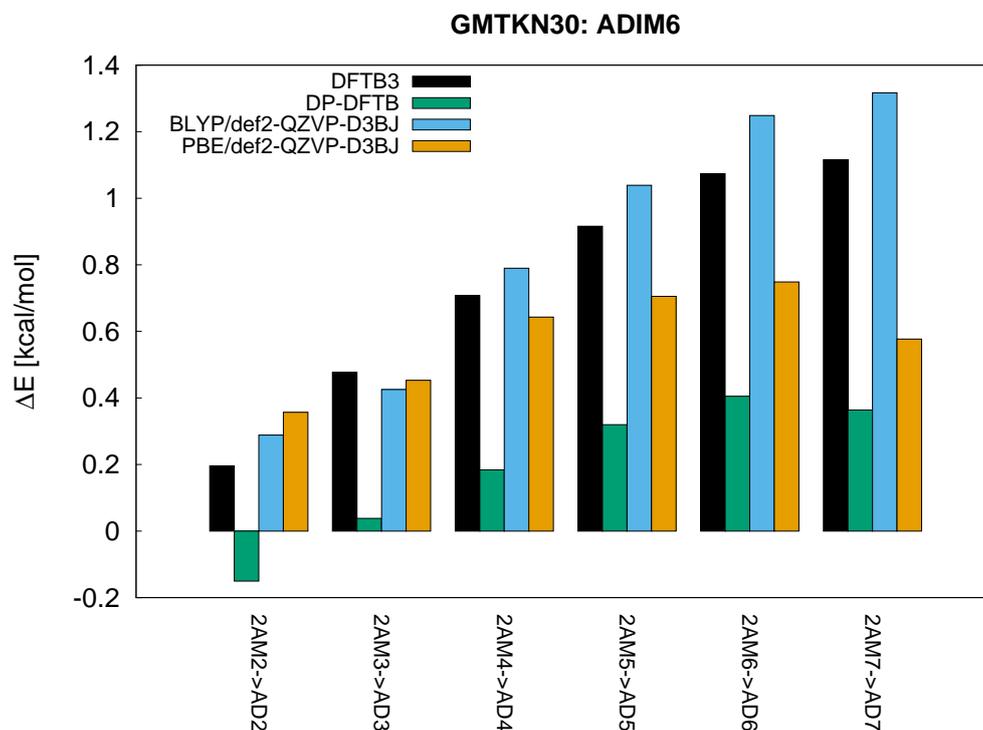


Figure 9.8: Deviation of DFTB-D3 and DP-DFTB from the CCSD(T) reference energies for the ADIM6 reactions. The overall MAD for DFTB-D3 amounts to 0.75 kcal/mol and is decreased with DP-DFTB to 0.24 kcal/mol. For comparison also the deviation of BLYP/def2-QZVP-D3BJ and PBE/def2-QZVP-D3BJ are listed.

IDISP

The IDISP is related to the ADIM6 set since it also contains London dispersion-dominated interaction energies. However, in the IDISP set these interactions occur intramolecular. Furthermore, the IDISP set contains with reactions like the anthracene dimerization, the hydrogenation of [2.2]paracyclophane etc. a more advanced CH-based chemistry compared to ADIM6 with its interactions of simple alkane chains. The performance of DFTB-D3, DP-DFTB, BLYP/def2QZVP-D3BJ and PBE/def2QZVP-D3BJ for this benchmark set is depicted in Figure 9.9. DP-DFTB yields for this subset no improvement compared to DFTB-D3. Instead, the MAD for DFTB-D3 is with 2.88 kcal/mol marginally lower than the one of DP-DFTB (3.28 kcal/mol).

Compared to the expensive DFT methods the tight-binding models perform very well and are able to outperform them (e.g. $\text{antdimer} \rightarrow \text{ant}$). Regarding the DFT methods, again the PBE functional performs in nearly ever case better than BLYP. Moreover, the $\text{antdimer} \rightarrow \text{ant}$ reaction somehow poses a problem to BLYP.

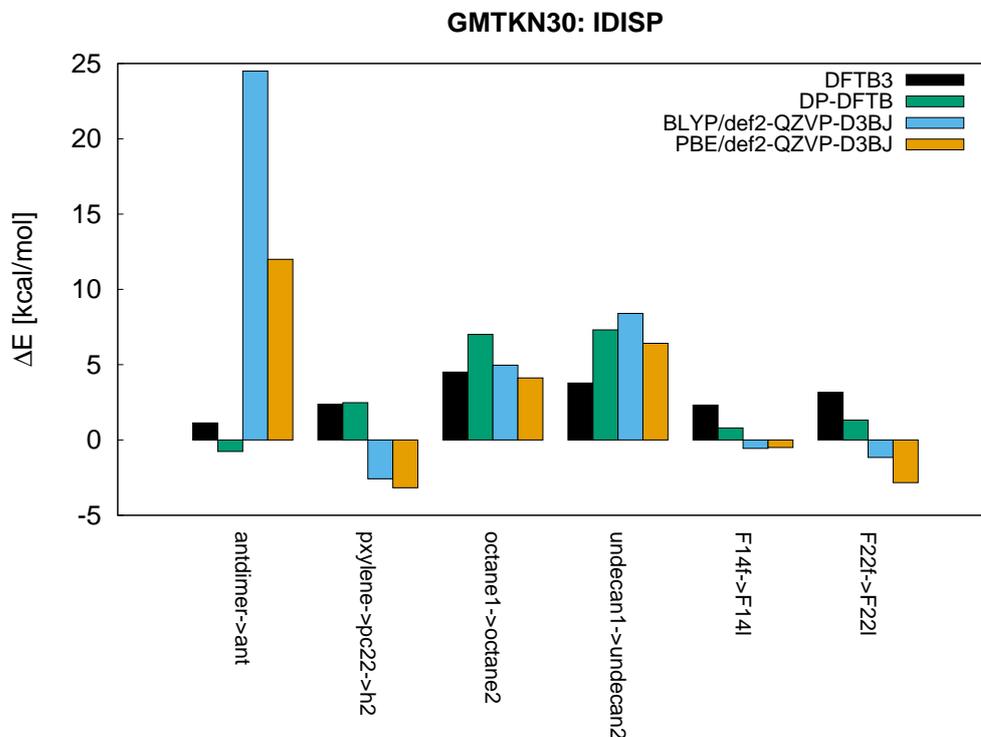


Figure 9.9: Deviation of DFTB-D3 and DP-DFTB from the reference energies for the IDISP reactions. The overall MAD for DFTB-D3 amounts to 2.88 kcal/mol and is increased with DP-DFTB to 3.28 kcal/mol. For comparison also the deviation of BLYP/def2-QZVP-D3BJ and PBE/def2-QZVP-D3BJ are listed.

L7

Figure 9.10 depicts the results for Hobza’s L7 set. This benchmark set is related to the IDISP and ADIM6 set since it comprises large dimer and trimer complexes that are stabilized through intermolecular dispersion interactions. Contrary to the IDISP and ADIM6 set, the L7 set is CHNO-based. Furthermore, most of the monomers exhibit (conjugated) π -systems like circumcoronene, guanine etc. which are strongly stabilized through π - π -stacking in the complex. A comparison of the DFTB-D3 results with the one of DP-DFTB reveals a similar performance of these methods like for the IDISP set: DFTB-D3 has a lower MAD with 1.92 kcal/mol than DP-DFTB with 2.83 kcal/mol. But again, with respect to the dissociation of the guaninetramer and octadecanedimer DFTB-D3 is not in every case the better tight-binding method.

The performance for the ADIM6, IDISP and L7 set depends mainly on the quality of the D3 parameters. Considering the ADIM6 results, the DP-DFTB D3 parameters yield a better description of alkane chains that are stabilized by dispersion, whereas the DFTB3 D3 parameters treat dispersion-stabilized aromatic systems of the IDISP/L7 set better.

Again, the expensive GGA density functional BLYP and hybrid density functional M06-2X perform not always better than DFTB-D3 or DP-DFTB.

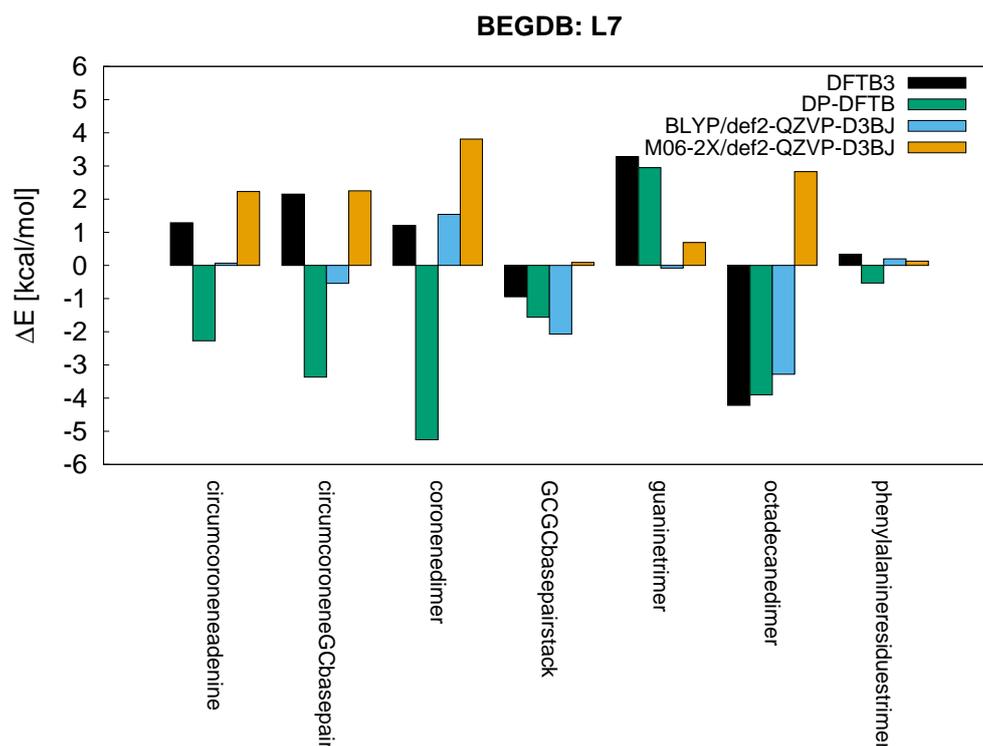


Figure 9.10: Deviation of DFTB-D3 and DP-DFTB from the CCSD(T) reference energies for the L7 reactions. The overall MAD for DFTB-D3 amounts to 1.92 kcal/mol and is increased with DP-DFTB to 2.83 kcal/mol. For comparison also the deviation of BLYP/def2-QZVP-D3BJ and M06-2X/def2-QZVP-D3BJ are listed.

SCONF

The SCONF benchmark set of Grimme's GMTKN30 database is made up of 15 D-galactitol (CX in Figure 9.11) and four D-glucopyranose (GX in Figure 9.11) conformers²⁰⁰. The Delta-Pauli correction is able to improve most of the conformational energies compared to DFTB-D3. The MAD amounts for DP-DFTB to 1.42 kcal/mol and for DFTB-D3 to 2.04 kcal/mol. For each conformation DFTB-D3 underestimates the barrier. Especially for the D-glucopyranose conformations the error of DFTB-D3 is large (-6/-7 kcal/mol). Since the energies of the underlying conformations depend mainly on rearrangements of the sugar ring (chair, boat, skew) and exocyclic groups²⁰¹ a correct description of the short-range repulsion is required. This is probably the case why DP-DFTB performs better, although still not perfectly.

BLYP and PBE show a decent performance for this benchmark set so that they mostly outperform the tight-binding models.

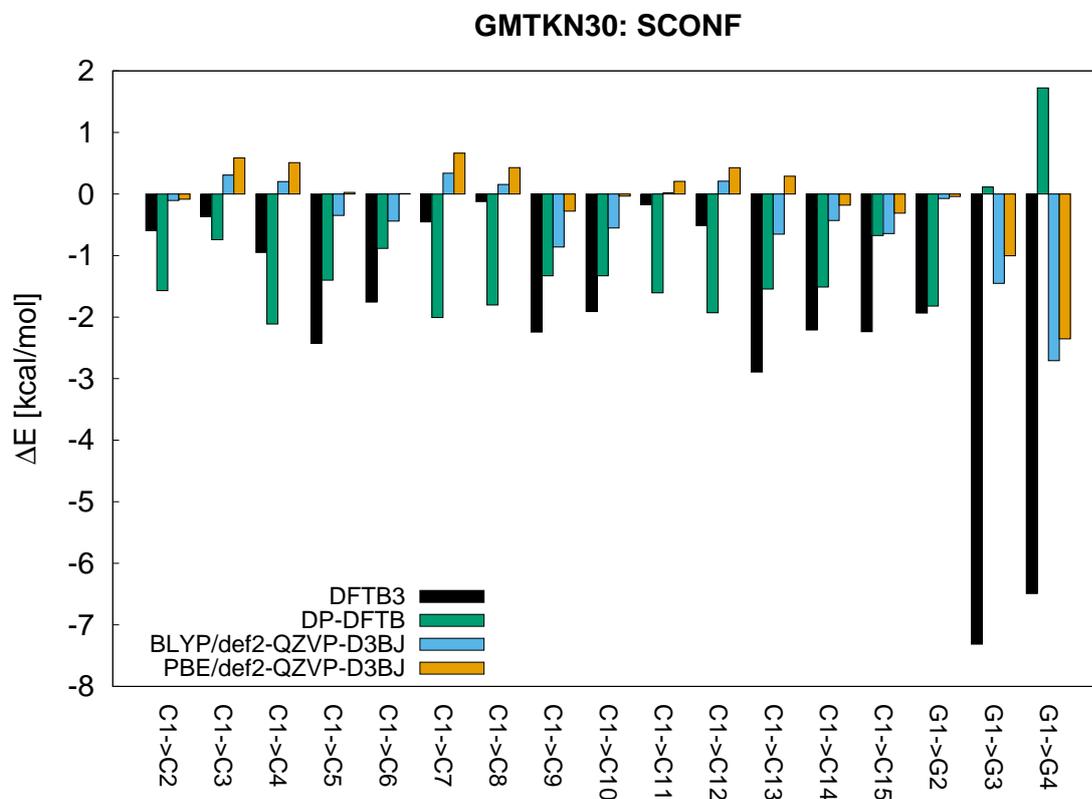


Figure 9.11: Deviation of DFTB-D3 and DP-DFTB from the CCSD(T) reference energies for the SCONF conformers. The overall MAD for DFTB-D3 amounts to 2.04 kcal/mol and is decreased with DP-DFTB to 1.42 kcal/mol. For comparison also the deviation of BLYP/def2-QZVP-D3BJ and PBE/def2-QZVP-D3BJ are listed.

FGG(PCONF)

The next small benchmark set covers peptide conformations. This so-called PCONF set was originally designed by Řeha et al.²⁰². It contains reactions of phenylalanyl-glycyl-glycine tripeptide conformations that are mainly stabilized via London dispersion interactions between the aromatic ring, the peptide bonds and the carboxylic acid. In 2008, Hobza's group extended this set and recalculated it to yield the FGG set¹⁹⁴. With respect to Figure 9.12, DFTB-D3 performs for this set with a MAD of 1.09 kcal/mol better than DP-DFTB (MAD=1.42 kcal/mol). Nevertheless, DP-DFTB outperforms in some cases DFTB-D3 (e.g. FGG99→FGG412 etc.).

Like for the other dispersion-dominated benchmark systems (ADIM6, IDISP, L7) the DFT methods do not perform much better than DFTB-D3 or DP-DFTB. This leads to the assumption that for a sufficient prediction of the FGG energies, probably the quality of the D3 parameters is crucial instead of full DFT and a large basis set.

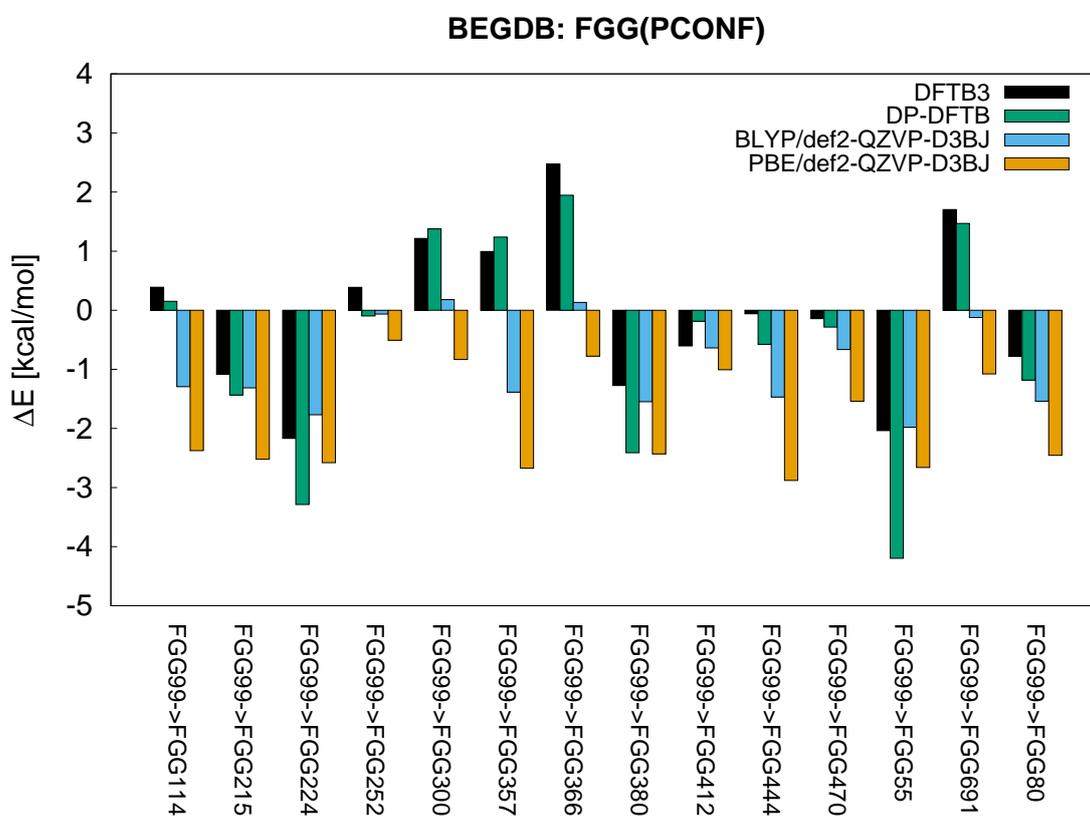


Figure 9.12: Deviation of DFTB-D3 and DP-DFTB from the CCSD(T) reference energies for the SCONF reactions. The overall MAD for DFTB-D3 amounts to 1.09 kcal/mol and is increased with DP-DFTB to 1.42 kcal/mol. For comparison also the deviation of BLYP/def2-QZVP-D3BJ and PBE/def2-QZVP-D3BJ are listed.

WATER27

The WATER27 subset is a benchmark set composed of dissociation energies of small to large neutral, protonated and deprotonated water clusters designed by Bryantsev et al.²⁰³. Recently, Manna et al.¹⁹³ released revised energies for this benchmark set at the MP2/CCSD(T) level which we used for comparison. The MAD analysis for this benchmark set yielded for the Pauli repulsion corrected DFTB method a MAD of 5.44 kcal/mol so that it outperforms DFTB-D3 with its MAD of 5.91 kcal/mol. However, a closer look at the bar chart in Figure 9.13 reveals that DP-DFTB mainly improves the dissociation energies of neutral and deprotonated water clusters, whereas DFTB-D3 performs better for protonated water clusters.

With respect to the expensive DFT functionals, the tight-binding models are able to outperform them especially for neutral water clusters. Moreover, the PBE functional suffers from a systematic error increase with growing system size. This is not the case for BLYP. Furthermore, the DFT methods suffer from a systematic barrier-overestimation.

Since dispersion interactions are negligible for the underlying reactions, the improvement of DP-DFTB over DFTB-D3 should result from the increased short-range Pauli-repulsion.

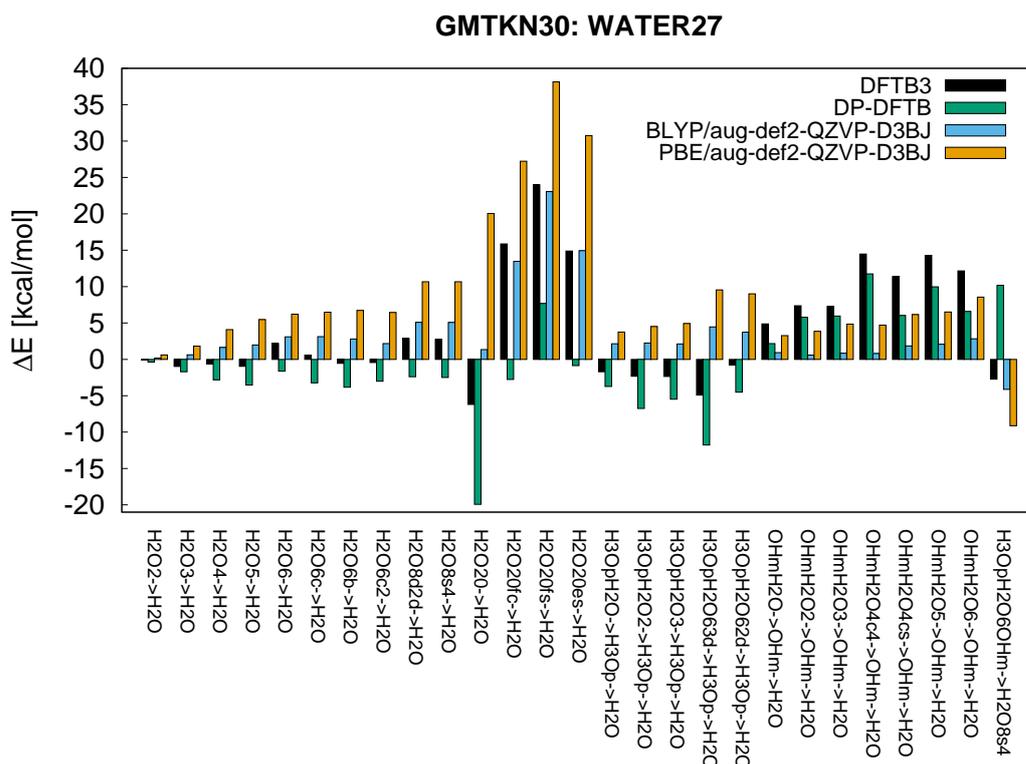


Figure 9.13: Deviation of DFTB-D3 and DP-DFTB from the CCSD(T)/MP2 reference energies for the WATER27 reactions. The overall MAD for DFTB-D3 amounts to 5.91 kcal/mol and is decreased with DP-DFTB to 5.44 kcal/mol. For comparison also the deviation of BLYP/def2-QZVP-D3BJ and PBE/def2-QZVP-D3BJ are listed.

ϕ/ψ -Potential Energy Surface of Alanine Dipeptide

Since Rosskyt's first MD simulation of a solvated alanine dipeptide in the 1970s²⁰⁴, this system has evolved over the years to "the benchmark system for peptides and proteins". Not only for quantum-chemical but also for force-field based methods²⁰⁵. This predestination as peptide benchmark is justified due to two main features of the system: Alanine dipeptide is the smallest molecule that exhibits peptide-like ϕ/ψ angles and thus is able to resemble the conformation of a peptide backbone. Secondly, the study of the backbone angles remains concise since the system yields only one Ramachandran plot⁹². In vacuo, six main conformations are found on the alanine dipeptide PES (see Figure 9.14D). There are three minima which are stabilized via intramolecular hydrogen bonds: the C_7^{eq} , C_7^{ax} and C_5^{ext} conformation. C_7^{eq} represents the global minimum. According to MP2/aug-cc-pVQZ data⁹², the second lowest minimum on this surface is C_5^{ext} followed by C_7^{ax} . These conformations exhibit an internal hydrogen bond between the carboxy oxygen and amino hydrogen atom. In the C_7 conformations this internal hydrogen bond builds together with the peptide backbone a seven-membered ring, whereas it is five-membered in the C_5^{ext} conformation. The difference between the C_7 conformations is the orientation of the methyl side chain: In C_7^{eq} it lies in the ring plane and in C_7^{ax} it stands perpendicular to it. The higher-energy conformers (β -turn, β_2 , right- and left-handed α -helix) exhibit no hydrogen bonds.

Figure 9.14A-C depicts the alanine dipeptide PES for BLYP, DFTB-D3 and DP-DFTB respectively. The BLYP-optimized geometries were taken from Christensen et al.⁹² and reevaluated with DFTB-D3 and DP-DFTB. The global minimum structure in the C_7^{eq} area was arbitrarily set to zero. Let us discuss the differences of the PES between the BLYP reference and the two tight-binding models. Considering the $C_7^{\text{eq}} \rightarrow C_5^{\text{ext}}$ transition, the barrier height amounts in DFTB-D3 and DP-DFTB to approximately 2.5 kcal/mol and is in accordance with the DFT reference. The difference between DFTB-D3 and DP-DFTB lies in the conformational energy of the C_5^{ext} structure: In DFTB-D3 it is 0.5 kcal/mol too low, whereas DP-DFTB is able to reproduce the reference conformational energy of 2.0 kcal/mol. The neighboring non-minimum conformations, β_2 and α , form in the BLYP representation an isoenergetical surface. In both tight-binding approaches this property is missing. Let us continue with the $C_7^{\text{eq}} \rightarrow C_7^{\text{ax}}$ transition: The MEP for this process leads over the α conformation in the BLYP reference and has a barrier of approximately 8.5 kcal/mol. Both DFTB methods are able to reproduce this feature and propose a MEP over the α conformation. DFTB-D3 underestimates the pathway over α by ca. 1.5 kcal/mol, whereas DP-DFTB is able to reduce this deviation to ca. 1.0 kcal/mol. Accordingly, DP-DFTB made a step in the right direction. On the other hand, the C_7^{ax} minimum is with DP-DFTB 0.5 kcal/mol lower than in DFTB-D3 and hence would be the second lowest minimum on the ϕ/ψ PES which is not correct regarding the BLYP reference. For the last conformer transition – the $C_7^{\text{ax}} \rightarrow C_5^{\text{ext}}$ transition – the barrier is underestimated with DFTB-D3 by ca. 2.5 kcal/mol compared to the BLYP reference. DP-DFTB is able to minimize this deviation such that the barrier only differs by ca. 2.0 kcal/mol.

The previous results have shown that DP-DFTB is able to introduce systematic improvements to the ϕ/ψ PES of alanine dipeptide such that the overall BLYP reference MAD for this surface is reduced slightly from 1.09 kcal/mol with DFTB-D3 to 0.85 kcal/mol. Since London dispersion interactions should play a minor role for this polar system, the improvement of DP-DFTB results probably from the auxiliary Pauli repulsion. The only drawback of DP-DFTB is that C_7^{ax} is too low in energy.

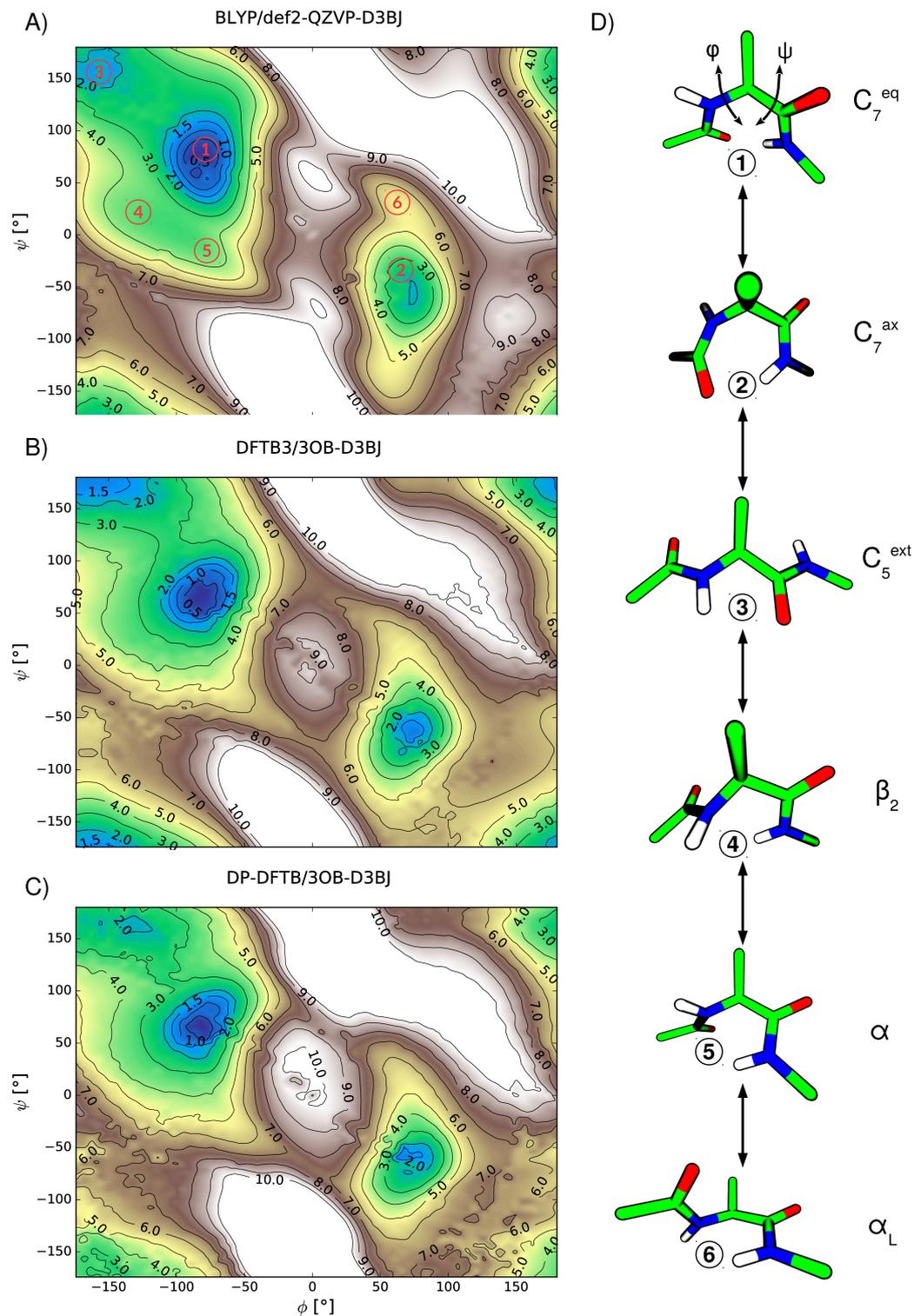


Figure 9.14: ϕ/ψ PES of alanine dipeptide in vacuo evaluated with BLYP/def2QZVP-D3BJ (A), DFTB3/3OB-D3BJ (B) and DP-DFTB/3OB-D3BJ (C). The energetical landscape exhibits six minimum and significant conformations that are depicted in D. For the sake of clarity, the conformations in D have been numbered from one to six and their position on the PES is shown in A.

Sugar Pucker Conformations of Deoxy-/Ribonucleotides

The function of carbohydrates in nature is like for proteins very widespread. In deoxy-/ribonucleotides, the building blocks of deoxyribonucleic acid (DNA)/ribonucleic acid (RNA), the ribose takes on different conformations. This effect results from the reduction of steric and electronic repulsion within the ring. A conformational change in the sugar ring will also affect the phosphate backbone and the nucleobase and hence the global conformation of the nucleic acid (A-DNA, B-DNA, etc.)¹⁹⁷. Thus, a correct description of these biological components is desirable. It has become good practice to benchmark the performance of semiempirical quantum and also force field methods on sugar puckered systems⁹². The standard peptide-specialized forcefields but also semiempirical quantum-methods mostly fail in the correct description of these systems. As a result, carbohydrate-specific forcefields like GLYCAM06²⁰⁶ and semiempirical methods that were fitted against carbohydrates like PM3CARB-1²⁰⁷, AM1/d-CB1²⁰⁸ and a B-spline corrected DFTB3¹⁹⁷ have evolved.

The here reported DFTB3-D3 vs. DP-DFTB benchmark treats the following (deoxy)ribonucleosides: deoxyadenosine (dA), deoxyguanosine (dG), deoxycytidine (dC), deoxythymidine (dT), adenosine (rA), guanosine (rG), cytidine (rC) and uridine (rU). For which, MP2-optimized geometries from Huang et al.¹⁹⁷ were used. In order to distinguish between the different sugar pucker conformations, two coordinates Z_x and Z_y are introduced which are connected to the two endocyclic torsions ν_1 and ν_3 (see Figure 9.15D) by:

$$Z_x = \frac{\nu_1 + \nu_3}{2 \cos(4\pi/5)} \quad Z_y = \frac{\nu_1 - \nu_3}{2 \cos(4\pi/5)} \quad (9.1)$$

The PES in space of Z_x/Z_y for the DNA nucleoside deoxyadenosine (dA) is depicted in Figure 9.15. Regarding the reference calculations, the global minimum on this surface is formed by the C2'-endo conformation (known as "western" minimum), and a secondary minimum, the C3'-endo conformation (known as "eastern" minimum) is ca. 3 kcal/mol higher in energy. Both basins are connected via two transition states. The so-called "northern" TS O4'-endo lies with 4.4 kcal/mol marginally lower than the "southern" TS O4'-exo with 4.9 kcal/mol. Figure 9.15B represents the DFTB-D3 results for this PES. Compared to the reference, there are three main discrepancies, which are also observed for the nucleosides dC, dG and dT (see Figure B.4-B.6, p.144):

1. The two minima basins are nearly isoenergetic. For the nucleoside dC, this trend even gets worse insofar the C3'-endo conformation becomes the global minimum (see Figure B.4B).
2. The barrier heights are underestimated by ca. 3.5 kcal/mol.
3. The preference for the transition states is interchanged in the DFTB-D3 representation (particularly observed in the PES of dC) so that the path which leads over the "southern" TS (O4'-exo) is lower in energy than via the "northern" TS (O4'-endo).

Now let us evaluate if and to what extent DP-DFTB can fix the listed weaknesses of DFTB-D3. Since the underlying systems are very polar, dispersion interactions probably play a minor role, and an improvement should mainly result from increased Pauli repulsion. The PES obtained with DP-DFTB for dA is depicted in Figure 9.15C. The plot shows clearly that the C3'-endo conformation is raised in the DP-DFTB representation by approximately 1 kcal/mol without affecting the location of the C2'-endo conformation which is good. Nevertheless, this conformation is still too low in energy by roughly 2 kcal/mol. The Delta-Pauli correction also resolves the wrong ordering of the minima which were observed for the nucleoside deoxycytidine. Furthermore, Figure 9.15C illustrates that DP-DFTB makes a step in the right direction in order to raise the barrier heights between the minima: The "southern" barrier is raised from ca. 0.6 kcal/mol with DFTB-D3 to ca. 1.6 kcal/mol with DP-DFTB. The same trend could be observed for the "northern" TS O4'-endo: The barrier amounts to ca. 1.2 kcal/mol in DFTB-D3, whereas it is raised to ca. 2.6 kcal/mol with DP-DFTB. However, in comparison with the reference, DP-DFTB still underestimates these barriers by 2-3 kcal/mol. Up to now, two of the three stated issues have been solved with the Pauli repulsion corrected tight-binding method. Nevertheless, the Delta-Pauli correction is not able to repair the interchanged preference for the "northern" and "southern" TS. Huang et al.¹⁹⁷ mentioned that the "southern" TS is unfavorable due to the steric contact of the C4'-hydroxy group with the C1' attached nucleobase (see Figure 9.15D). In more detail, in the "southern" TS of the dA nucleoside the C4'-hydroxy group is directly facing a 3.05 Å distant CH-group of the nucleobase. Maybe a DP-DFTB parameter set with an increased repulsion for the atom pair C-O could also fix this issue. The above stated improvements also hold true for the other deoxynucleosides dC, dG and dT (see Figure B.4-B.6 in the Appendix).

Compared to DNA, the RNA-sugar pucker profiles in Figure B.7-B.10 (see Appendix) exhibit just one TS and the "eastern" and "western" minimum are closer in energy. The MP2 reference provides either C2'-endo or C3'-endo as global minimum, different case by case. In all DFTB3-D3-PES of the ribonucleosides only the "western" C2'-endo minimum is visible and in the case of rA, rC and rU it is with respect to the MP2 reference the non-global one. The corresponding DP-DFTB results for the ribonucleosides show that also the "eastern" minimum (and sometimes a TS) can be resolved. Nevertheless, these minima are weak and still too high in energy. Again, DP-DFTB is able to improve the barrier heights also for the ribonucleoside puckering profiles.

An interesting and problematic observation in all PES calculated with DP-DFTB is that large negative values of Z_y are unfavorable (see for example Z_y region from 0° to -20° in Figure 9.15C). The corresponding MP2 and DFTB-D3 plots are much broader in this region (0° – -40°), meaning the potential energy is too steep in the DP-DFTB representation. A decrease of the Z_y parameter mainly results in a decrease of the distance between the O4' oxygen atom and the nitrogen atom of the sugar puckers N-glycosidic bond. This finding is an evidence for an overestimated N-O-repulsion in the Delta-Pauli model.

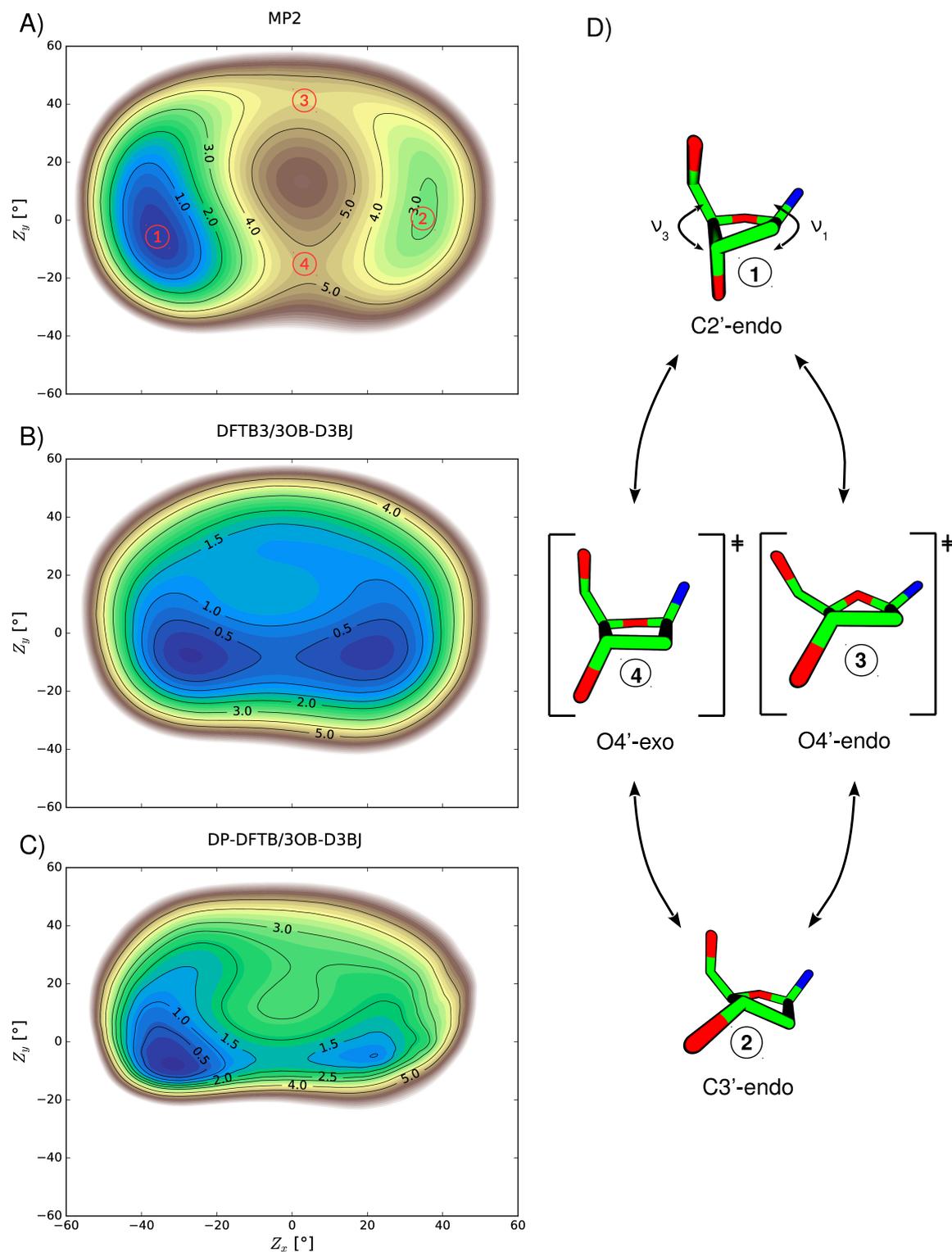


Figure 9.15: PES for deoxyadenosine (dA) defined by the two puckering parameters Z_x and Z_y . The reference profile calculated with MP2/6-311++G(3df,2p)//MP2/6-31++G(d,p) is shown in A. The corresponding PES calculated with DFTB3/3OB-D3BJ and DP-DFTB/3OB-D3BJ is shown in B and C respectively. The individual conformations that are sampled by these two pseudorotations are depicted in D.

9.5 Excursus: Long-Range Corrected DFTB2 for Non-Covalent Interactions

A general drawback of conventional LDA-/GGA-based DFT functionals is the insufficient description of long-range electron-electron exchange interactions. However, a proper description of these interactions is crucial for a sufficient prediction of e.g. valence orbital energies, electron excitation spectra, response properties or van der Waals interactions²⁰⁹. Global hybrids like B3LYP or PBE0, which incorporate a specific amount of HF exchange, are known to improve the description of electronic and vibrational properties²¹⁰.

The positive effect of the admixture of HF exchange is also exploited in long-range corrected (LC) DFT functionals. In contrast to global hybrid functionals, in LC functionals the short-range part of the exchange functional is combined with the long-range part of the HF exchange:

$$E_{xc} = E_{xc}^{\text{DFT},\omega} + E_x^{\text{HF},\omega} \quad (9.2)$$

with $E_{xc}^{\text{DFT},\omega}$ as the short-range DFT electron exchange-correlation energy, $E_x^{\text{HF},\omega}$ as the long-range HF exchange energy and ω as the range-separation parameter. In contrast to conventional DFT functionals, LC functionals provide the correct asymptotic $-1/r$ behavior of the Kohn-Sham potential, minimize the electron self-interaction error and the overdelocalization of electron-density^{211–217}.

A formalism for a long-range correction for DFTB2 was reported in 2012²¹⁸ and implemented in 2015²¹⁹. However, a change in the functional in DFTB is not as straightforward as in DFT since the semiempirical method relies on precomputed Hamiltonian matrix elements derived from DFT. Furthermore, a reparametrization of the electronic parameters also requires an adjustment of the repulsive potential. Recently, these reoptimized electronic and repulsive parameters were released by our group¹²³. In the course of reparametrization we augmented the LC-DFTB2 model with the D3BJ dispersion model in order to account for attractive London dispersion effects. This was achieved by a fitting approach comparable to the one used for Delta-Pauli: The D3 parameters a_1 , a_2 and s_8 were PSO-fitted against reference energies from the S66 set^{98,a} for the three LC-DFTB parameter sets *base*, *shift* and *split*. The fitted parameters are listed in Table 9.6. A detailed benchmark of the

Table 9.6: Final D3 dispersion parameters for the three LC-DFTB2 parameter sets.

Parameter Set	base	shift	split
a_1	0.717	0.816	0.497
a_2	2.565	2.057	3.622
s_8	0.011	0.010	0.010

newly parametrized LC-DFTB2-D3 is given in ref. 123. In this publication, it is mentioned that LC-DFTB2-D3 shows an improvement over DFTB2 for geometries and vibrational frequencies such that it performs similarly well as DFTB3/3OB. However, for atomization and reaction energies the method is less efficient than DFTB3/3OB.

^aNot the S66x8.

In this work, we focus on the performance of the derived D3 parameters. Figure C.1 in the Appendix shows that the D3 dispersion correction efficiently improves the prediction of S66 energies with LC-DFTB2^b. Consequently, the D3 correction reduces the LC-DFTB2 MAD from 2.4 kcal/mol to 0.7 kcal/mol. With this, the D3-corrected LC-DFTB2 method is also able to outperform the D3-corrected DFTB3 method (MAD=0.8 kcal/mol). However, on closer examination of Figure C.1 it becomes apparent that the D3 parameters mainly improve the energies of polar systems such that unpolar systems still exhibit a large error. This is surprising, since especially unpolar systems which are dispersion-dominated, should benefit from the inclusion of a dispersion correction. This observation can also be made in benchmark calculations of GMTKN30 subsystems. Figure 9.7 shows MAD values for the WATER27, FGG(PCONF) and IDISP sets calculated with different DFTB methods. Also, these results show that polar WATER27 systems benefit from the dispersion correction more than the unpolar systems in the FGG(PCONF) subset. For the heavily dispersion-dominated IDISP system, the inclusion of the D3 dispersion correction makes the results even worsen. This effect was also reported in ref. 123. It was suggested that the D3 parameters overcompensate the error of LC-DFTB2 for polar systems.

In order to further investigate this artifact, we took one polar system – the AcOH-AcOH dimer – and one unpolar system – the pentane-pentane dimer – from the S66x10^{220,c} set and recalculated the interaction energies with DFTB2/MIO, DFTB3/3OB-D3BJ, LC-DFTB2/BASE and LC-DFTB2/BASE-D3BJ. The results are depicted in Figure 9.16. Regarding the polar AcOH-AcOH dimer system all methods yield reasonable results and reproduce the trend of the CCSD(T) reference (see Figure 9.16A). The corresponding equilibrium geometries of the semiempirical methods are also in accordance with the reference (see Figure 9.16C).

However, for the unpolar pentane-pentane dimer system (see Figure 9.16B) the LC-DFTB2 methods show a physically wrong behavior in the short-range region ($r/r_0 = 0.7 - 0.95$). The most significant error of both LC methods is the shift of the equilibrium distance to a ratio of $r/r_0 = 0.8$. Moreover, both LC-DFTB2 methods show an enormous lack of repulsion in the short-range region. The two other tight-binding models, DFTB2 and DFTB3-D3BJ reproduce the trend of the reference. Certainly, this physically incorrect behavior of the LC-DFTB2 method also affects the prediction of the equilibrium distances (see Figure 9.16D) such that the distance between the pentane monomers is significantly reduced. Further, we recalculated the whole S66x8 set with DFTB2, DFTB3-D3BJ, LC-DFTB2/BASE and LC-DFTB2/BASE-D3BJ. The corresponding dissociation plots in the Appendix (see Figure C.2, p.152) show that the unphysical behavior of the two LC-DFTB2 methods can be observed in every unpolar system.

To conclude, the PSO-derived D3 parameters are not responsible for the bad description of unpolar systems. It is also not the case, like the authors suggested that the D3 parametrization compensates to a great extent for an error of LC-DFTB2 for polar systems. Instead, we think that probably something in the parametrization of the confinement

^bThe Figure in the Appendix shows the results for the *base* parameter set. The results for the *split* and *shift* parameter sets are not shown but show an identical trend.

^cIdentical to the S66x8 set but extended by three short-range data points.

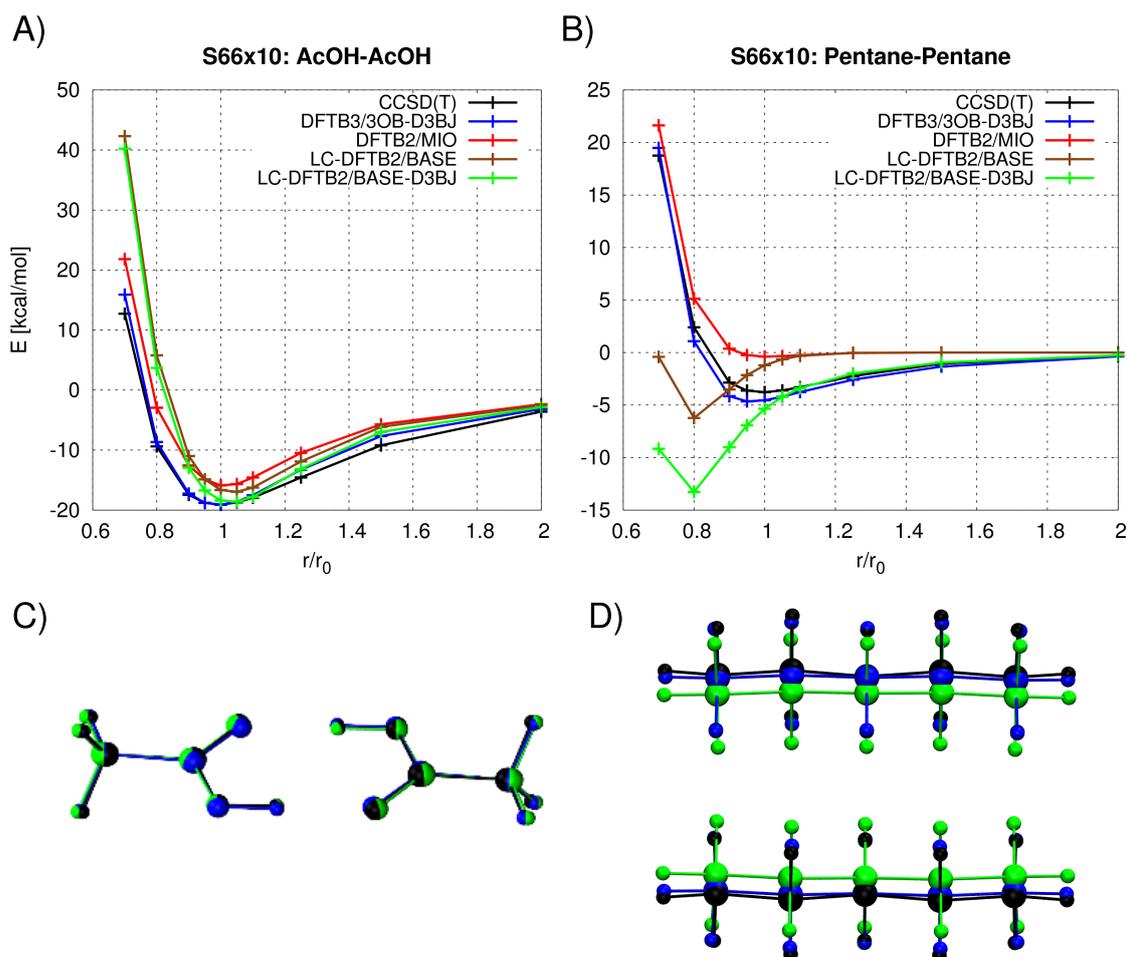


Figure 9.16: S66x10 dissociation curves and equilibrium geometries of the AcOH-AcOH and pentane-pentane dimer. A: All DFTB models are able to give a physically correct prediction of the polar AcOH-AcOH dimer dissociation curve. B: Also the corresponding equilibrium geometries of this system are predicted correctly with DFTB3/3OB-D3BJ (blue) and LC-DFTB2/BASE-D3BJ (green) compared to the CCSD(T) reference (black). C: For the unpolar Pentane-Pentane system the reparametrized LC-DFTB2 models (brown and green curve) are not able to give a physically correct prediction of the dissociation curve (especially the short-range part), while DFTB3/3OB-D3 and DFTB2/MIO can. D: This error of the LC-DFTB2 methods also affects the equilibrium distance. As a result, the distance of the pentane monomers is significantly reduced in the LC-DFTB2/BASE-D3BJ representation, while DFTB3/3OB-D3BJ is in accordance with the CCSD(T) reference.

radii of the C or H atom has gone wrong. This hypothesis is supported by the fact that already DFTB2/MIO yields reasonable results for the unpolar systems here presented (see Figure 9.7 and 9.16). Consequently, for future work these deficiencies have to be solved.

Table 9.7: Mean Average Deviations (in kcal/mol) of DFTB2/3 methods for selected subsets of the GMTKN30 database. The values after slash are calculated with D3-BJ dispersion included.

Subset	DFTB2/MIO	DFTB3/3OB	LC-DFTB2/base	LC-DFTB2/shift	LC-DFTB2/split
WATER27	22.0	7.1/5.3	19.9/12.6	19.9/12.1	20.0/13.5
FGG(PCONF)	1.9	2.0/1.1	1.2/ 1.0	1.4/ 1.0	1.7/ 1.0
IDISP	2.9	8.5/3.4	2.6/ 8.2	2.6/ 8.2	2.7/ 7.8

9.6 Conclusions

In this chapter a first parameter set for the empirical DFTB3 Delta-Pauli correction in combination with D3 dispersion was created. The parametrization comprised all combinations of CHNO atom pairs. The S66x8 and the newly created R10 set proved themselves as reasonable components of the PSO training set such that the description of short-range Pauli repulsion could be improved while maintaining the D3 performance for mid- to long-range dispersion. Accordingly, DP-DFTB shows improvements over DFTB-D3 regarding intramolecular reaction barriers (rotational barriers), short-range interaction energies and equilibrium geometries. These improvements are not unique to systems in the training set. The benchmark of Delta-Pauli on prominent test sets from databases like GMTKN30 or BEGDB, as well as on biomolecular systems of interest (alanine dipeptide, sugar pucker conformers), indicated a general improvement of Delta-Pauli compared to DFTB-D3. It is worth mentioning that the improvements of Delta-Pauli are not only of quantitative character. In some cases the empirical correction is able to convert a qualitatively wrong DFTB-D3 representation into a correct one like in the example of sugar pucker conformers.

However, systems which are heavily dominated by attractive dispersion (like IDISP) are probably better represented in the Pauli repulsion uncorrected DFTB3-D3 version. The results also showed that there is still too much repulsion on the O-N atom pair (cf. sugar pucker), whereas the C-C atom pair still lacks some repulsion (cf. n-butane). These observations hint at a certain imbalance in the CHNO training set. Furthermore, it was proven that Delta-Pauli is not applicable to all atom pairs. In order to avoid the disruption of hydrogen-bonded systems, the correction has to be deactivated for X-H atom pairs (where X is an electron-withdrawing group). The necessity for a special treatment of hydrogen-bonded systems also contributes to the fact that the parametrization of Delta-Pauli is not straightforward. This is aggravated by the large parameter space that the correction entails and the uncertainty of an unbalanced CHNO training set.

Nevertheless, the present results show that the missing short-range Pauli repulsion in DFTB may be introduced by applying an empirical correction to the DFTB total energy.

Summary and Outlook

The first part of this work addressed the mechanism of the O→bR transition in bacteriorhodopsin. Since structural information on the O and O* states is very sparse from experimental side, computational models for these and the bR ground state have been created by employing enhanced sampling MD simulations.

The results indicated that the O→bR transition entails structural rearrangements regarding the R82 side chain orientation and the PRG formation, which in turn control the hydration of a cavity between the D85/D212 and PRG site indirectly. Of particular interest in the O→bR transition was the investigation of the mechanism and free energy change along the assumed long-range PT. The modeled O state structures suggested that a prerequisite for this reaction is already given, since they exhibited water wires connecting the D85 site with the PRG. In order to achieve the aforementioned objectives, DFTB3-based QM/MM free energy simulations comprising Umbrella Sampling or Metadynamics have proven themselves as successful approaches. According to these calculations, the O→O* subprocess is a direct PT from D85 to D212 which involves only small structural rearrangements like the perturbation of the pentagonal HBN near D85, D212 and K216RET. Besides the direct PT, also less favorable, indirect pathways were observed where the proton is transferred via bridging water molecules in the form of hydronium or hydroxide. Counter-intuitively, the O state showed a preference of the hydroxide pathway over the hydronium one.

This observation also applies to the subsequent O*→bR transition, where the proton is transferred via an inverse Grotthuss/proton hole mechanism over a water wire from D212 to E204. Thus, the active site between D85/D212 and the PRG shows a general preference for the proton hole pathway. We assume that this effect results from the presence of the positively charged R82 amino acid in the PT pathway as well as of the electrostatic environment in the active site region. In contrast to O→O*, it was shown that the PT in the O*→bR transition is accompanied with two considerable structural rearrangements: the side chain reorientation of R82 and the PRG formation.

Like literature suggests, the calculated free energy profile for the O→bR transition indicated that the O→O* subprocess represents the rate-limiting step⁶⁴ and is furthermore an endergonic process⁶⁷. The subsequent subprocess O*→bR exhibits a 50% lower barrier

than the $O \rightarrow O^*$ transition and ensures the exergonicity of the $O \rightarrow bR$ transition. With this, the O^* state and hence the protonated D212, could be identified as a metastable state. However, with respect to the rate-limiting step, the $O \rightarrow bR$ transition would occur in the low μs regime, which deviates from the experimental estimate of 5 ms. A further discrepancy is the predicted exergonicity for the $O \rightarrow bR$ reaction which is too low with respect to experimental expectations.

For future work, we want to eliminate possible error sources in the complex free energy calculation of the $O^* \rightarrow bR$ subprocess. In order to ensure equilibration along the PT path, all structural processes beside the PT need to be equilibrated. Therefore multidimensional Umbrella Sampling or Metadynamics will be used. This approach should also yield further information about the timing of the PT-parallel side processes.

In the course of the PT studies on bR, the performance of the semiempirical DFTB3 method for this type of reaction was also benchmarked against the higher level theory method B3LYP. The results indicated a decent agreement of DFTB3 with B3LYP with a small reaction barrier underestimation.

In the second part of this work it was shown that the consistent underestimation of inter- and intramolecular reaction barriers in DFTB3 is caused by the lack of short-range Pauli repulsion. The description of this QM phenomenon may be introduced via the empirical correction model Delta-Pauli. In this work, the first parameter set for the Delta-Pauli model was derived for CHNO-based systems. This procedure also entailed the refitting of D3 dispersion parameters. Delta-Pauli benchmark calculations of selected organic and biochemical systems of interest showed overall promising results for properties where a decent description of Pauli repulsion is essential. In this way, especially the description of rotational barriers, non-covalent interaction energies at short distances but also equilibrium geometries of molecules showed an overall improvement over DFTB3. It is also worth mentioning that due to the refit of D3 parameters the Delta-Pauli model retains a good description of attractive mid- to long-range non-covalent interactions. With this, the Delta-Pauli model represents a paradigm how computationally efficient semiempirical quantum methods can be extended in order to yield results with a quality comparable to high level quantum chemistry methods (e.g. DFT with triple- ζ basis set).

Nevertheless, there are still some issues of Delta-Pauli which should be addressed in future work. For example, the results indicated that some atom pairs gain too less/ too much repulsion. This in turn is probably a consequence of an imbalanced training set, which is however hard to detect. The best approach to solve this issue, is to employ an optimization technique which automatically determines appropriate atom pair weights (e.g. like in ref. 124). Furthermore, in order to allow efficient MD simulations employing Delta-Pauli DFTB, analytical gradients have to be derived and implemented. It is also of interest to know, how Delta-Pauli describes macromolecular properties like for example the radial distribution function of water. Therefore periodic boundary conditions need to be implemented in Delta-Pauli.

Part IV

Appendix

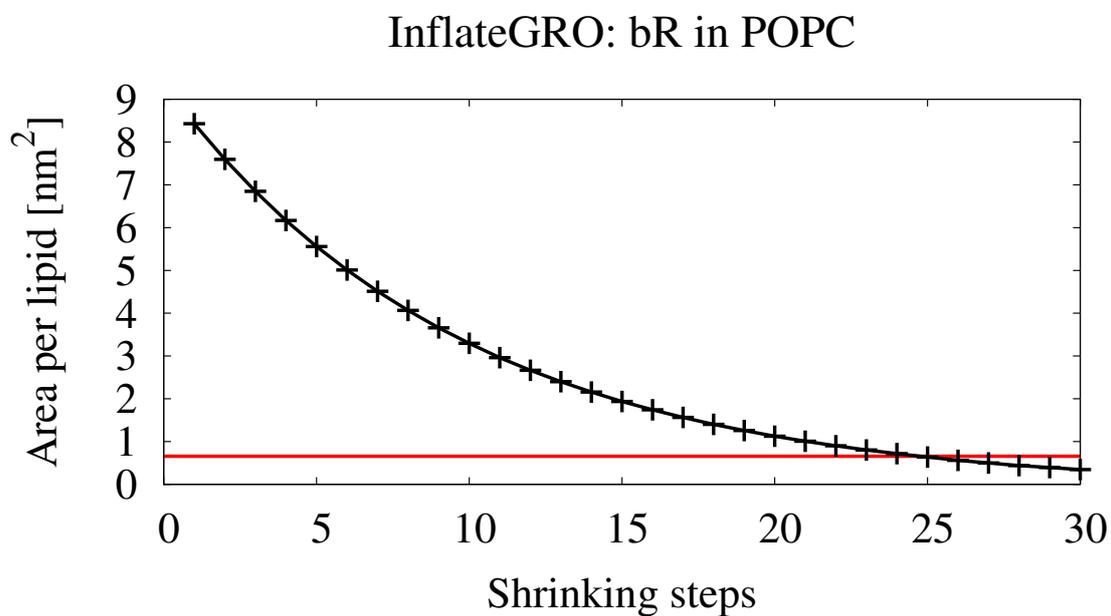
PT Reactions in bR

Figure A.1: Area per lipid ratio in nm² as a function of the InflateGRO shrinking steps. The experimental value of 65.8Å² is reached after approximately 25 steps.

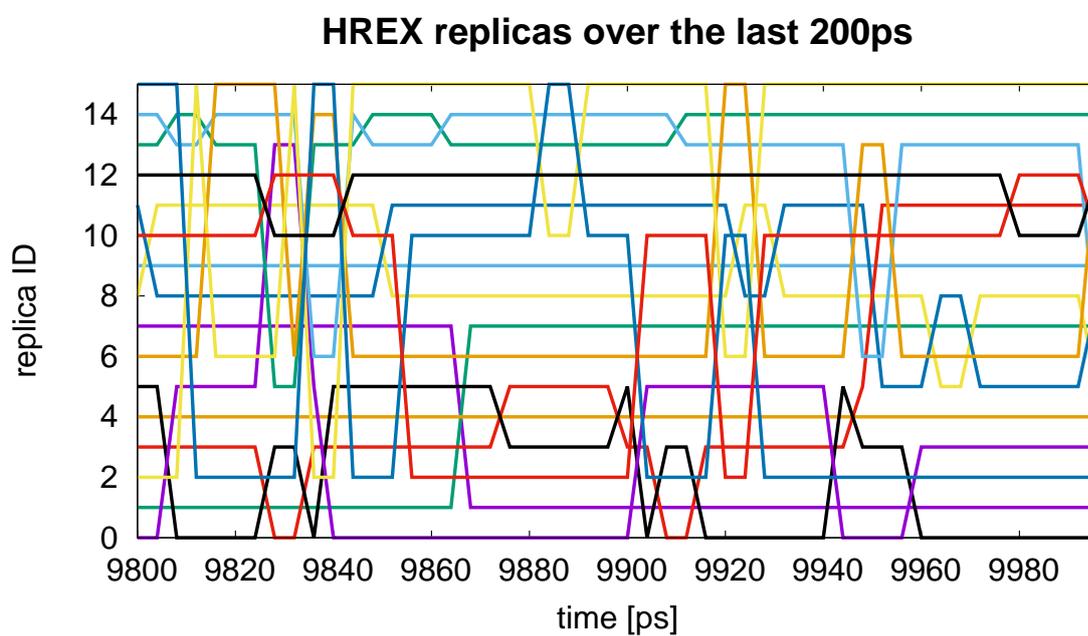


Figure A.2: Visualization of the exchange steps between the 16 replicas during the last 200 ps of HREX simulation. The data was taken from the bR state simulation.

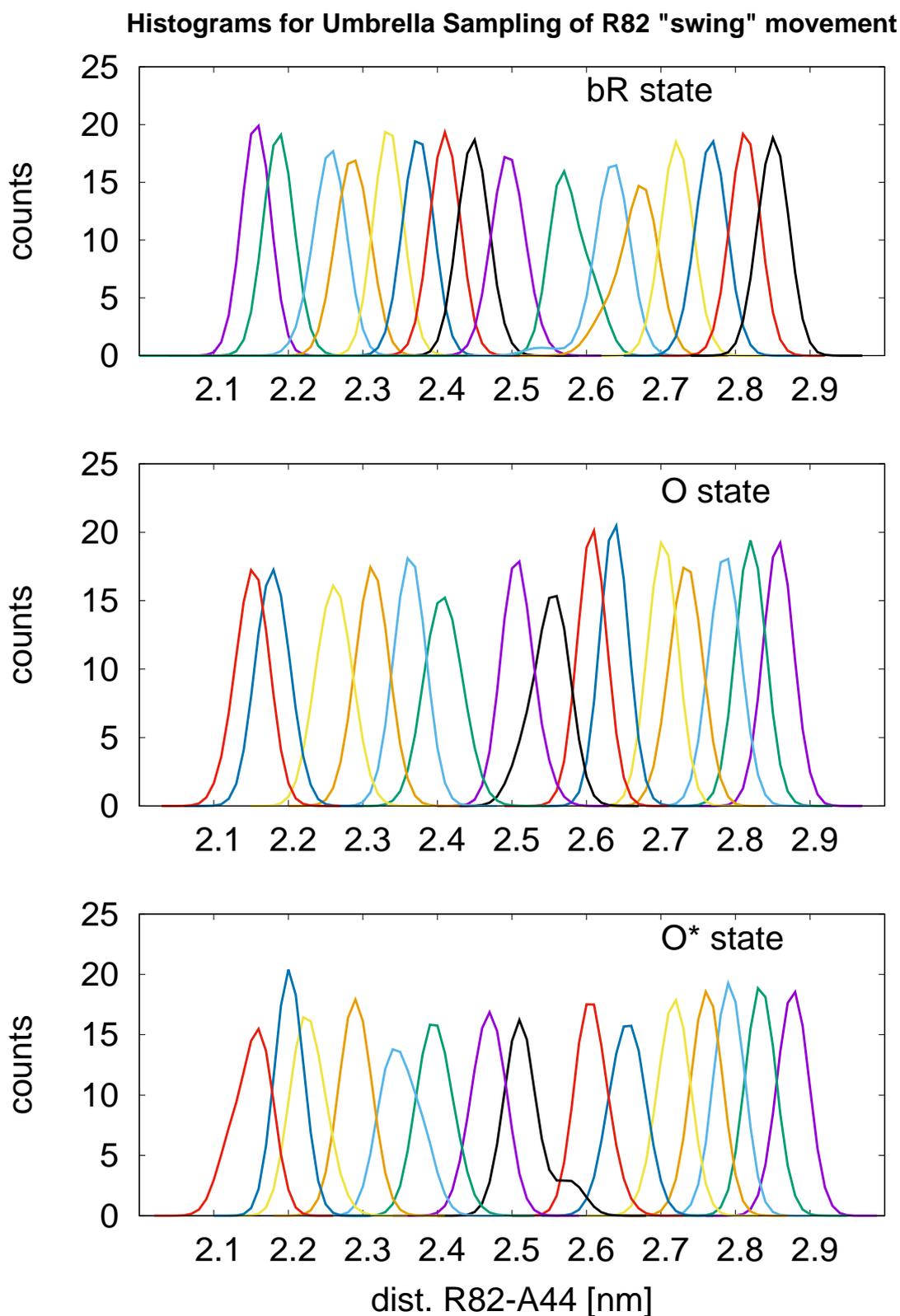


Figure A.3: The histograms of the Umbrella Sampling for the R82 swing movement show a reasonable overlap for all three bR protonation models.

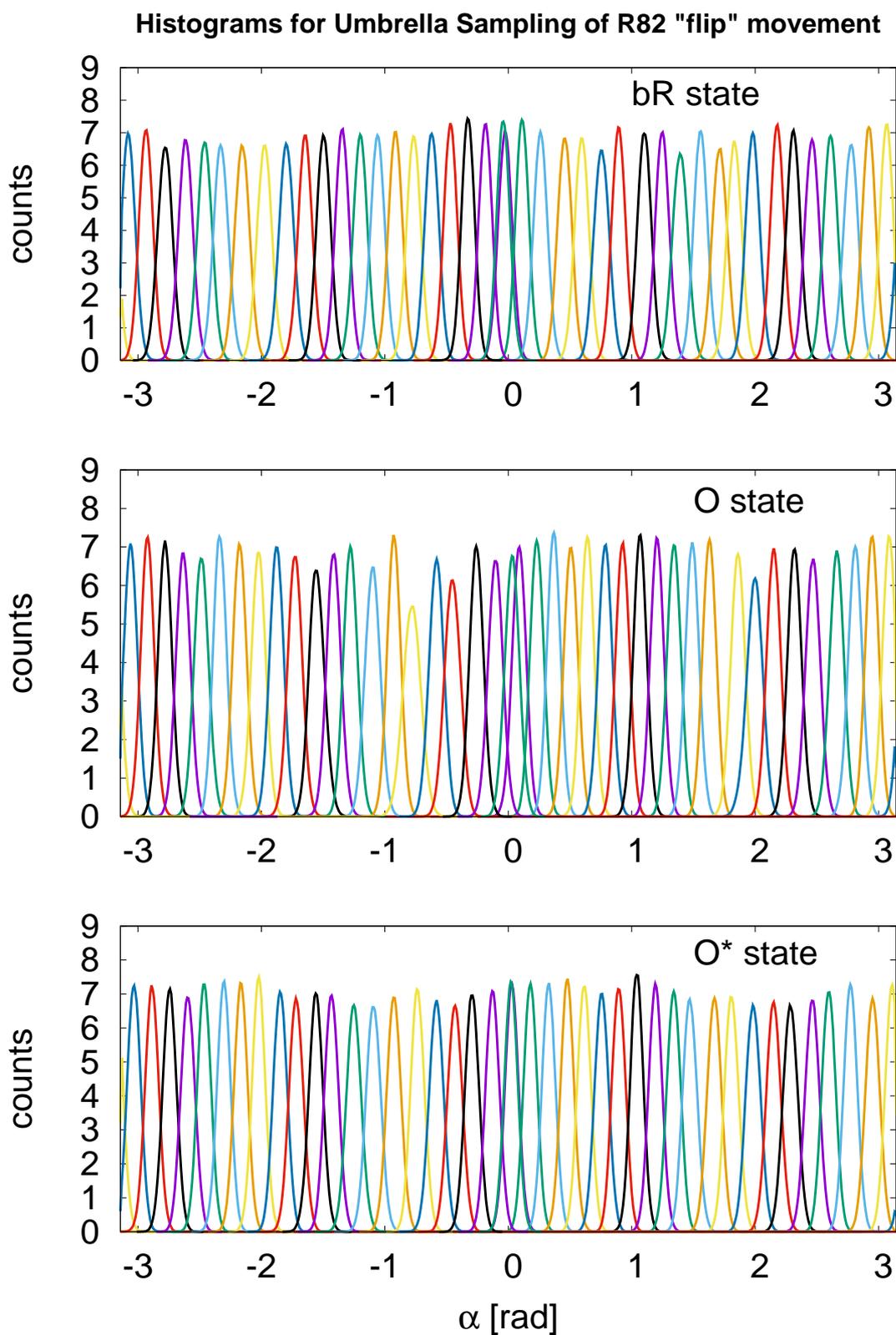


Figure A.4: The histograms of the Umbrella Sampling for the R82 flip movement show a reasonable overlap for all three bR protonation models.

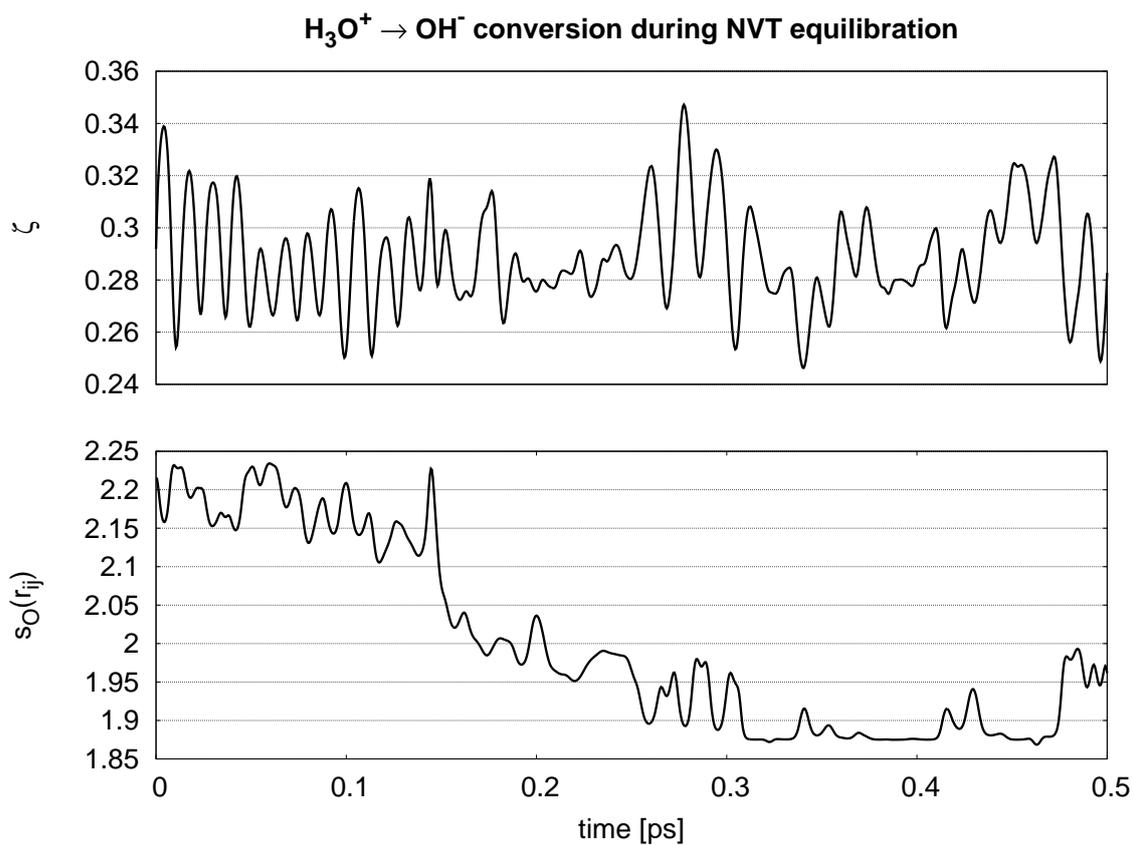


Figure A.5: During NVT equilibration the water chain is converted after approximately 150 fs from a hydronium-exhibiting to a proton-hole-exhibiting conformation. The change in conformation is only indicated by the $s_O(r_{ij})$ function and not by ζ .

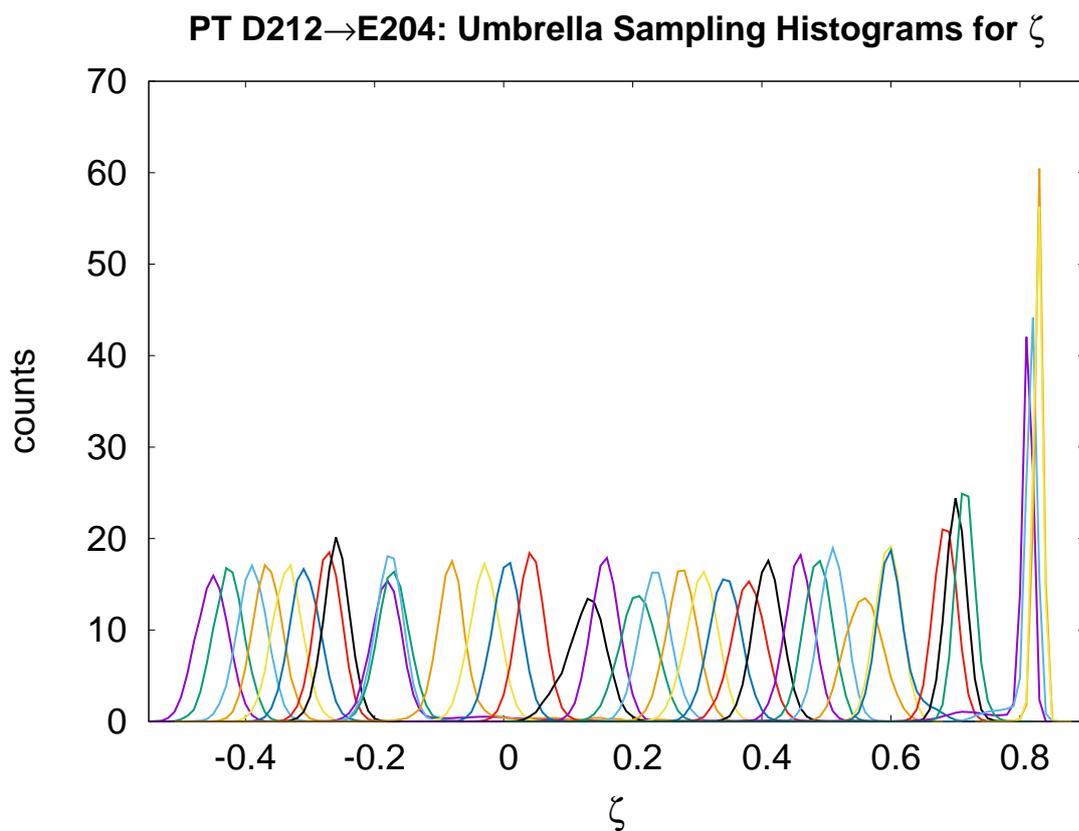


Figure A.6: ζ histograms for the Umbrella Sampling of the PT reaction from D212 to E204. The windows show sufficient overlap.

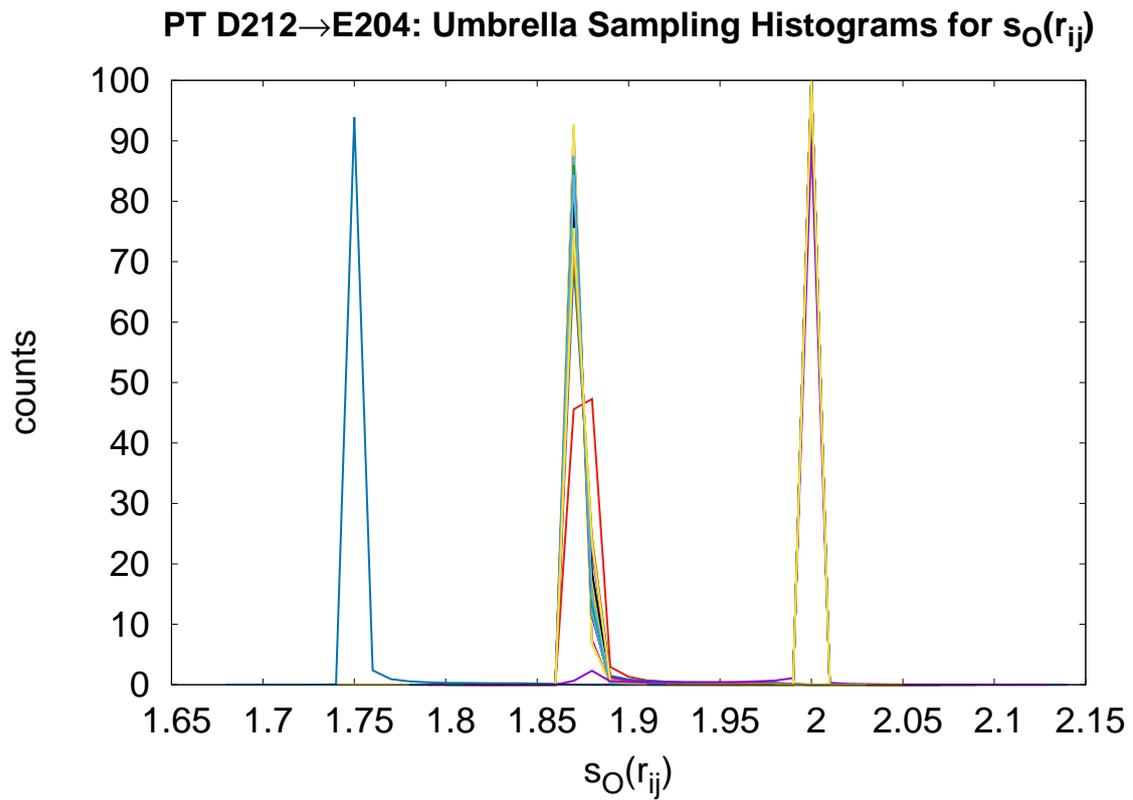


Figure A.7: $\overline{s_O(r_{ij})}$ histograms for the Umbrella Sampling of the PT reaction from D212 to E204. The windows show either a non-protonated water wire ($\overline{s_O(r_{ij})} = 2$) or a proton hole configuration of it ($\overline{s_O(r_{ij})} < 2$).

Listing A.1: Plumed input file for the 3D-MW-Metadynamics run.

```

1 # Initial PT donor/acceptor
2 dcom: COM ATOMS=1280
3 DON: POSITION ATOM=dcom NOPBC
4 acom: COM ATOMS=3292
5 ACC: POSITION ATOM=acom NOPBC
6
7 # Get positions from all atoms which are involved in the PT
8 OD1: POSITION ATOM=1281 NOPBC
9 OD2: POSITION ATOM=1282 NOPBC
10 HD: POSITION ATOM=1283 NOPBC
11 O1: POSITION ATOM=3684 NOPBC
12 H11: POSITION ATOM=3685 NOPBC
13 H12: POSITION ATOM=3686 NOPBC
14 O2: POSITION ATOM=3687 NOPBC
15 H21: POSITION ATOM=3688 NOPBC
16 H22: POSITION ATOM=3689 NOPBC
17 O3: POSITION ATOM=3690 NOPBC
18 H31: POSITION ATOM=3691 NOPBC
19 H32: POSITION ATOM=3692 NOPBC
20 O4: POSITION ATOM=3693 NOPBC
21 H41: POSITION ATOM=3694 NOPBC
22 H42: POSITION ATOM=3695 NOPBC
23 OA1: POSITION ATOM=3293 NOPBC
24 OA2: POSITION ATOM=3294 NOPBC
25
26 # Definition of the switching function components (3rd term of xi)
27 cx: COORDINATION_X GROUPA=1281,1282,3684,3687,3690,3693,3293,3294 GROUPB
    =1283,3685,3686,3688,3689,3691,3692,3694,3695 R_SW=0.115 D_SW=0.0045
28 cy: COORDINATION_Y GROUPA=1281,1282,3684,3687,3690,3693,3293,3294 GROUPB
    =1283,3685,3686,3688,3689,3691,3692,3694,3695 R_SW=0.115 D_SW=0.0045
29 cz: COORDINATION_Z GROUPA=1281,1282,3684,3687,3690,3693,3293,3294 GROUPB
    =1283,3685,3686,3688,3689,3691,3692,3694,3695 R_SW=0.115 D_SW=0.0045
30
31 # x-dimension of xi
32 COMBINE ...
33 LABEL=x
34 ARG=HD.x,O1.x,H11.x,H12.x,O2.x,H21.x,H22.x,O3.x,H31.x,H32.x,O4.x,H41.x,H42.x,cx
35 COEFFICIENTS=1,-2,1,1,-2,1,1,-2,1,1,-2,1,1,-1
36 PERIODIC=NO
37 ... COMBINE
38
39 #y-dimension of xi
40 COMBINE ...
41 LABEL=y
42 ARG=HD.y,O1.y,H11.y,H12.y,O2.y,H21.y,H22.y,O3.y,H31.y,H32.y,O4.y,H41.y,H42.y,cy
43 COEFFICIENTS=1,-2,1,1,-2,1,1,-2,1,1,-2,1,1,-1
44 PERIODIC=NO
45 ... COMBINE
46
47 #z-dimension of xi
48 COMBINE ...
49 LABEL=z
50 ARG=HD.z,O1.z,H11.z,H12.z,O2.z,H21.z,H22.z,O3.z,H31.z,H32.z,O4.z,H41.z,H42.z,cz
51 COEFFICIENTS=1,-2,1,1,-2,1,1,-2,1,1,-2,1,1,-1
52 PERIODIC=NO
53 ... COMBINE
54
55 # Define zeta
56 MATHEVAL ...

```

```

57 LABEL=zeta
58 ARG=x,y,z,DON.x,DON.y,DON.z,ACC.x,ACC.y,ACC.z
59 VAR=x,y,z,dx,dy,dz,ax,ay,az
60 FUNC=(sqrt((x-dx)*(x-dx)+(y-dy)*(y-dy)+(z-dz)*(z-dz))-sqrt((x-ax)*(x-ax)+(y-ay)*(y-ay)+(z-az)*(z-az)))/
(sqrt((x-dx)*(x-dx)+(y-dy)*(y-dy)+(z-dz)*(z-dz))+sqrt((x-ax)*(x-ax)+(y-ay)*(y-ay)+(z-az)*(z-az)))
61 PERIODIC=NO
62 ... MATHEVAL
63
64 # Distance between D85Cgamma-D212Cgamma)
65 dist: DISTANCE ATOMS=1280,3292
66
67 # Upper harmonic restraint
68 UPPER_WALLS ARG=dist AT=0.66 KAPPA=200000 LABEL=uwall
69
70 # Average oxygen coordination number of all QM-treated H2O molecules
71 coord: COORDINATIONNUMBER SPECIESA=3684,3687,3690,3693 SPECIESB
      =1283,3685,3686,3688,3689,3691,3692,3694,3695 R_0=0.12 NN=45 MM=90 MEAN
72
73 # Setup Metadynamics simulation
74 METAD ...
75 LABEL=metad
76 ARG=zeta,dist,coord.mean
77 SIGMA=0.065,0.02,0.025
78 GRID_MIN=-1.0,0.0,1.4
79 GRID_MAX=1.0,0.71,2.6
80 HEIGHT=0.2
81 PACE=200
82 BIASFACTOR=8
83 TEMP=300
84 WALKERS_N=20
85 WALKERS_ID=IDOFWALKER
86 WALKERS_DIR=../
87 WALKERS_RSTRIDE=500
88 ... METAD
89
90 # Print output
91 PRINT ARG=zeta,dist,coord.mean,metad.bias FILE=colvar STRIDE=10

```

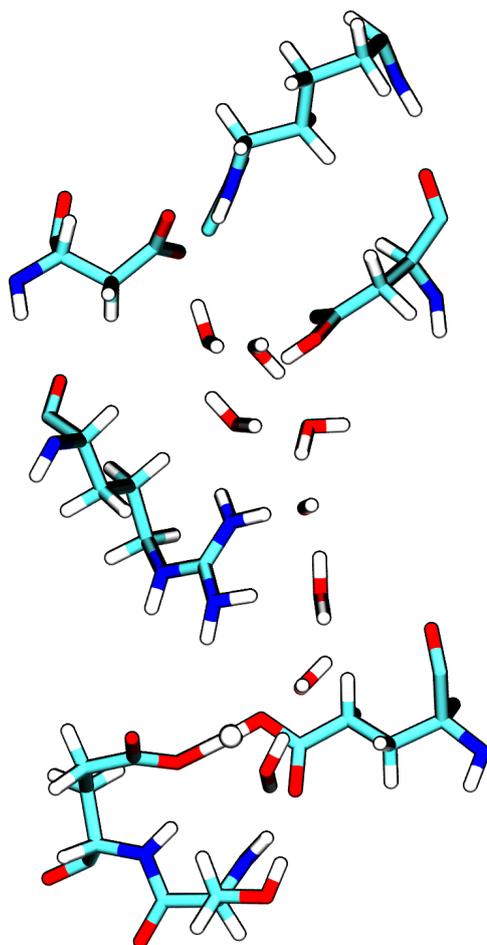


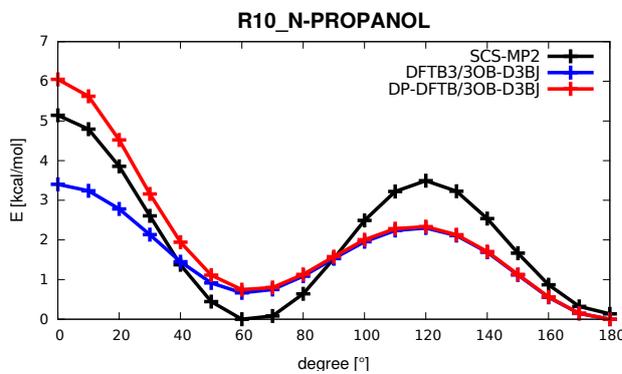
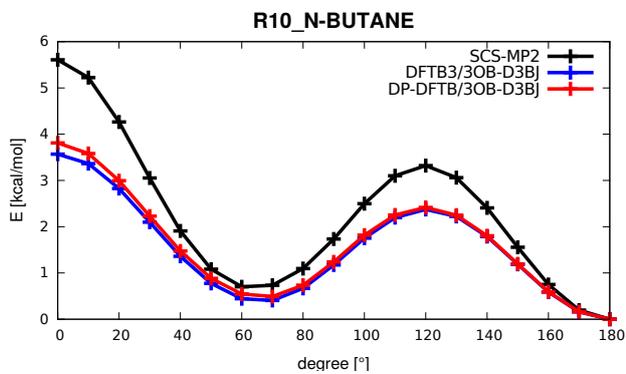
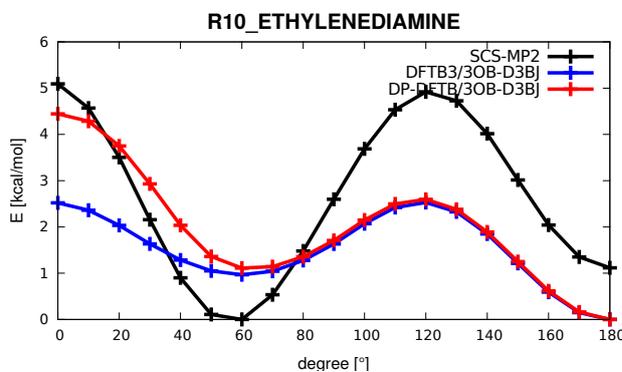
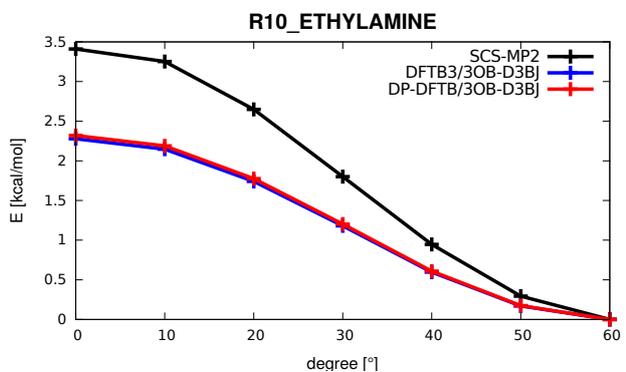
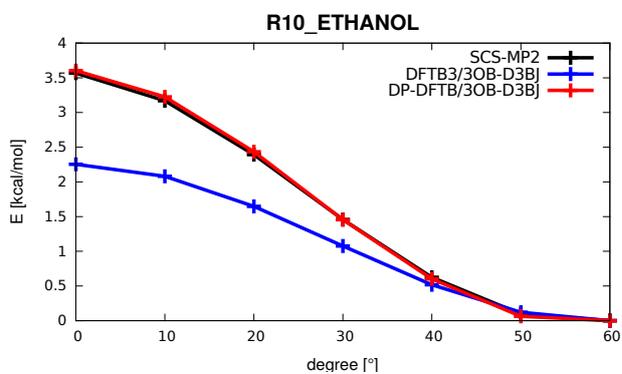
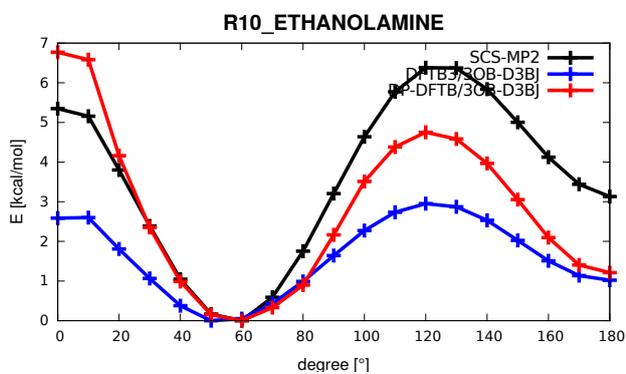
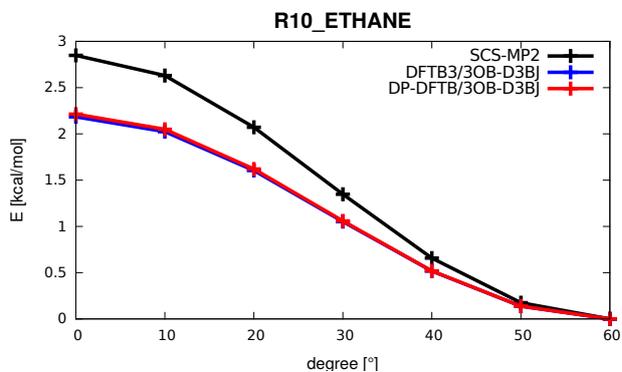
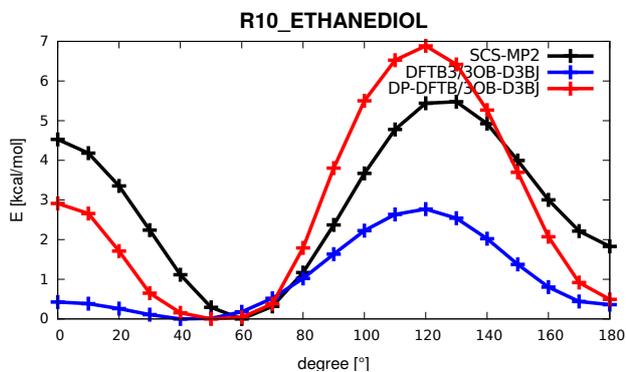
Figure A.8: Intermediate structure from the D212→E204 PT. The CEC is localized at $\zeta = 0$ and gets in this configuration stabilized by the R82 guanidinium group.

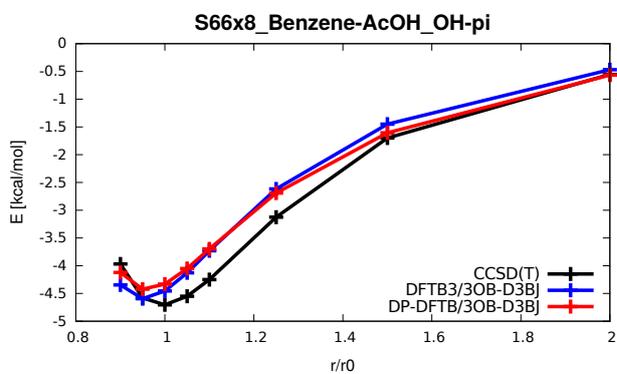
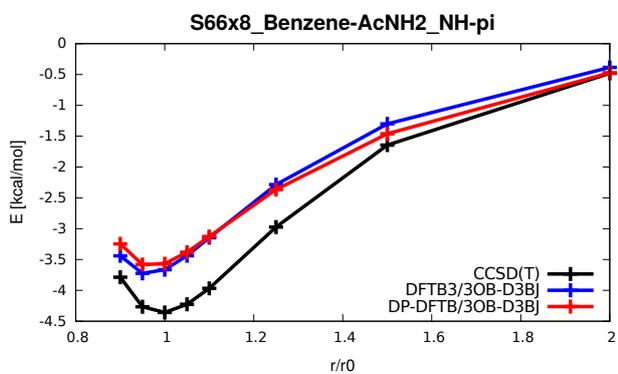
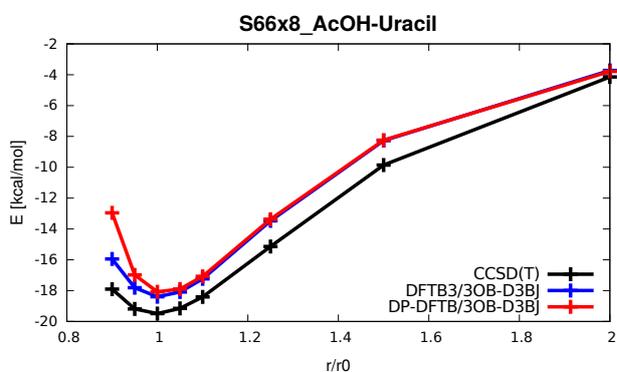
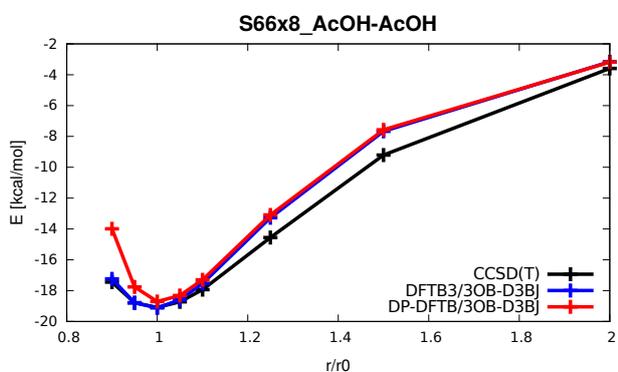
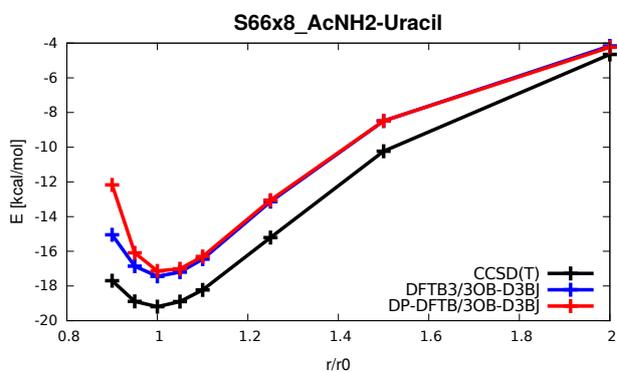
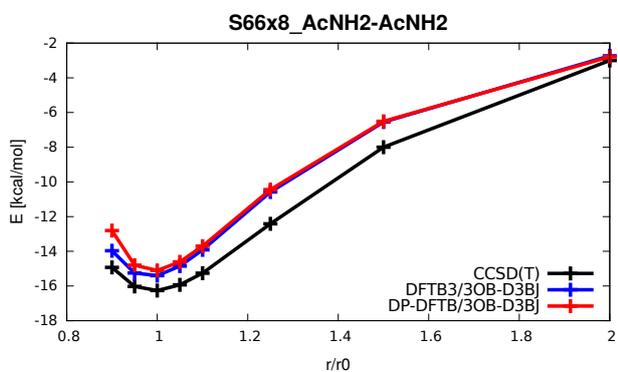
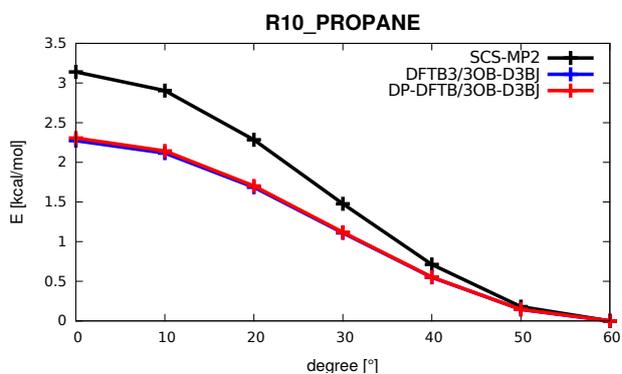
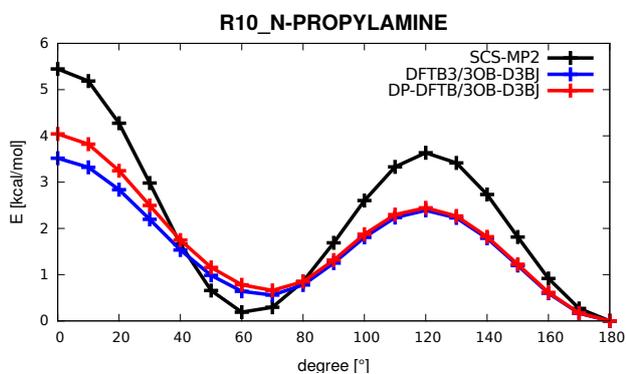
Delta-Pauli DFTB**Table B.1:** PSO constants used for fitting. With N = number of total particles, ω = inertia weight, ϕ_p = cognitive acceleration and ϕ_g = social acceleration.

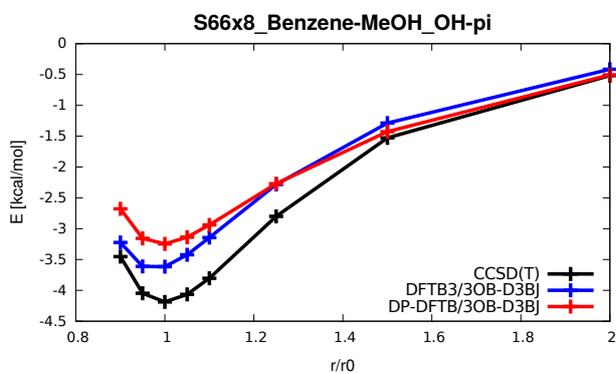
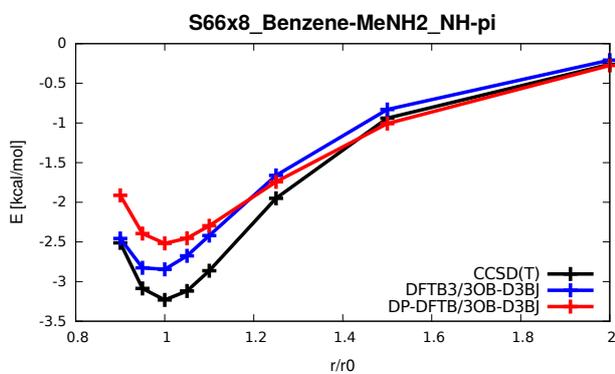
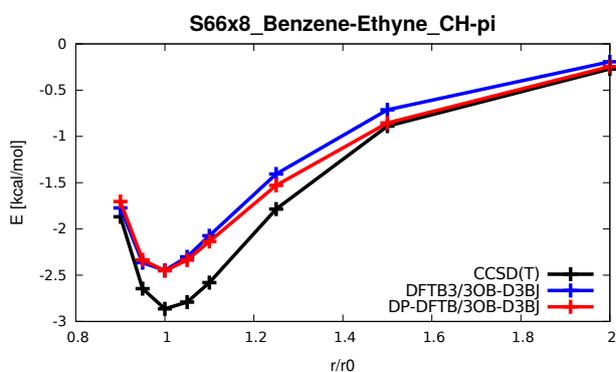
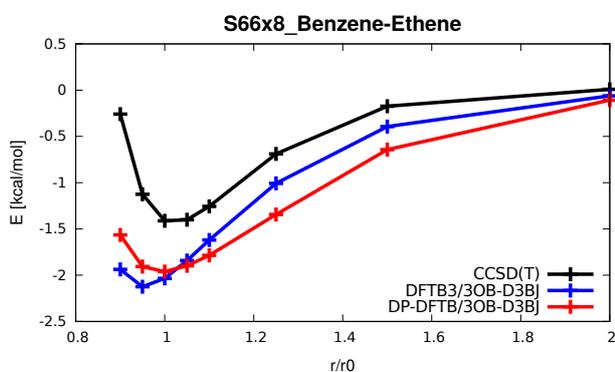
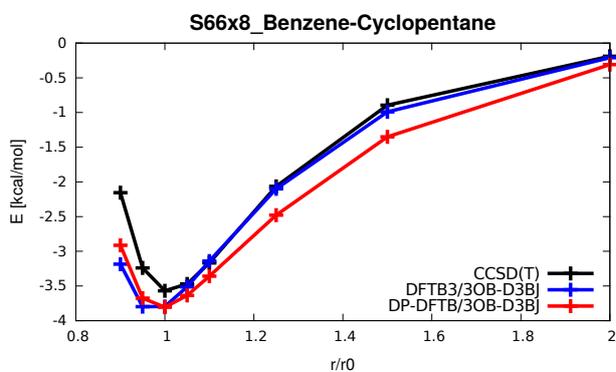
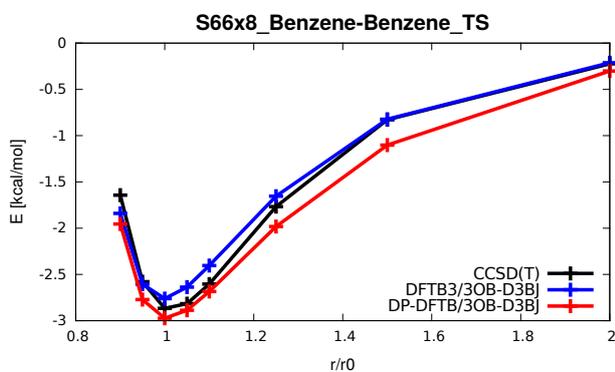
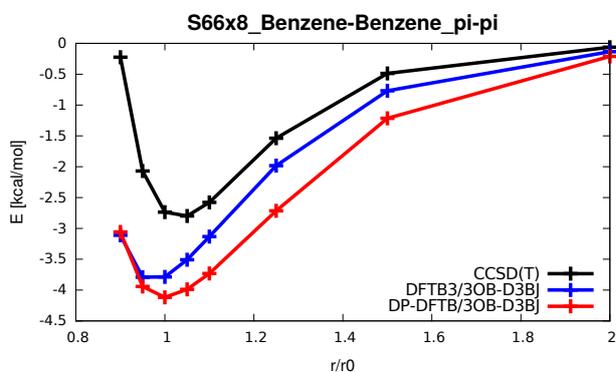
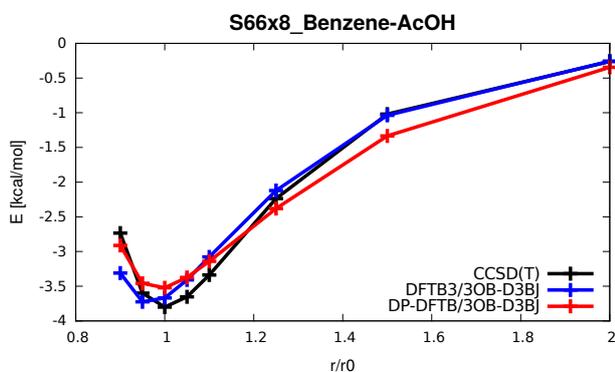
PSO constant	Value
N	204
ω	-0.2134
ϕ_p	-0.3344
ϕ_g	2.3259

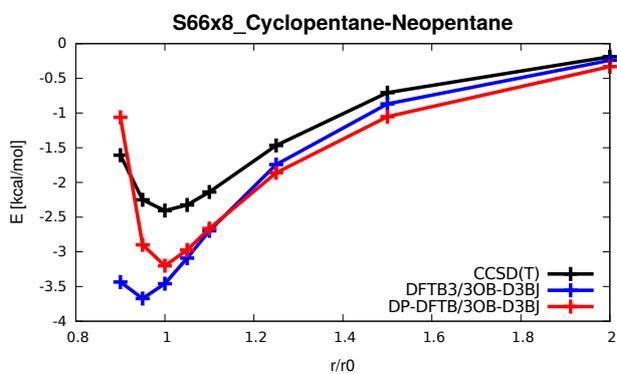
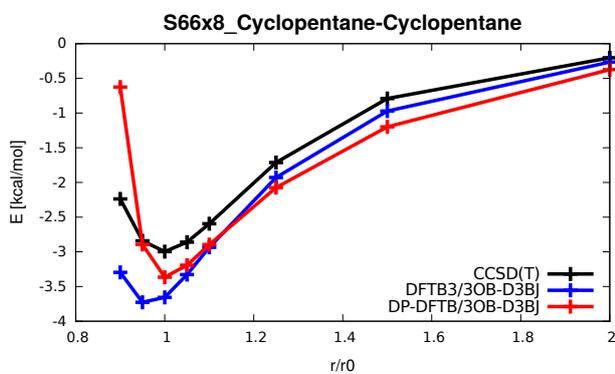
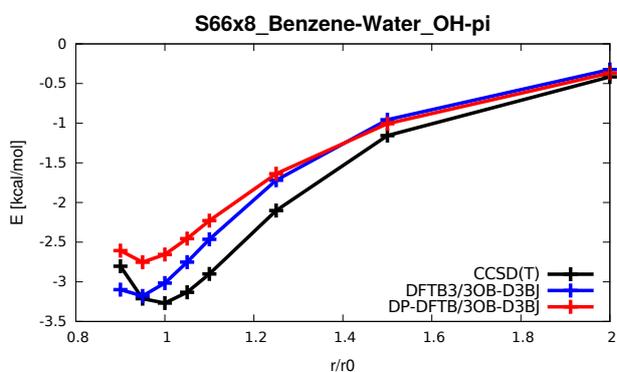
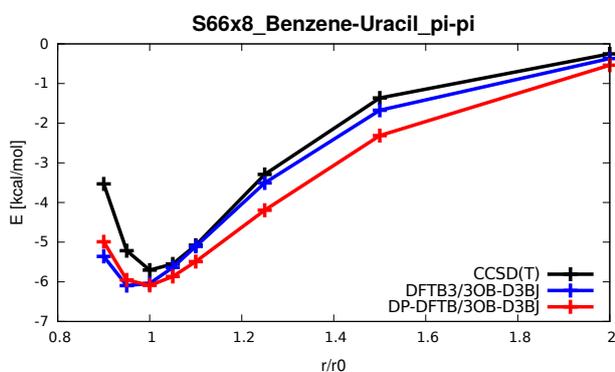
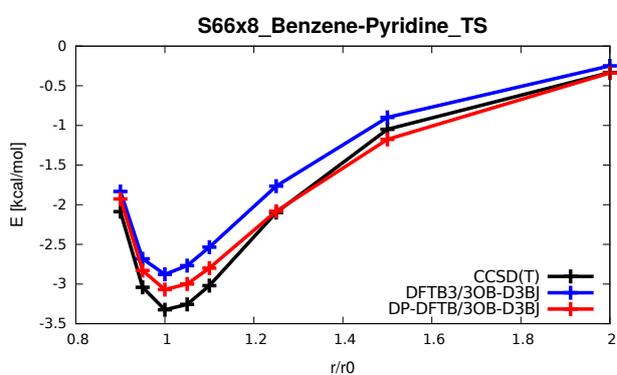
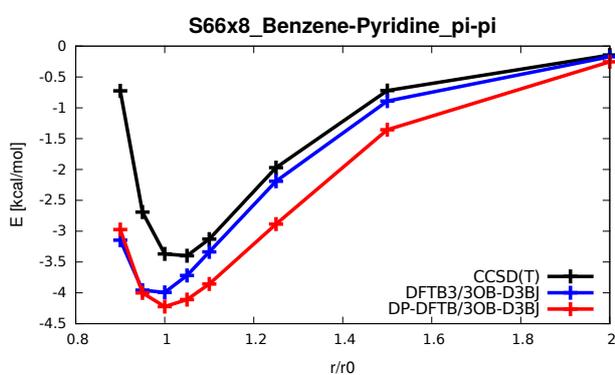
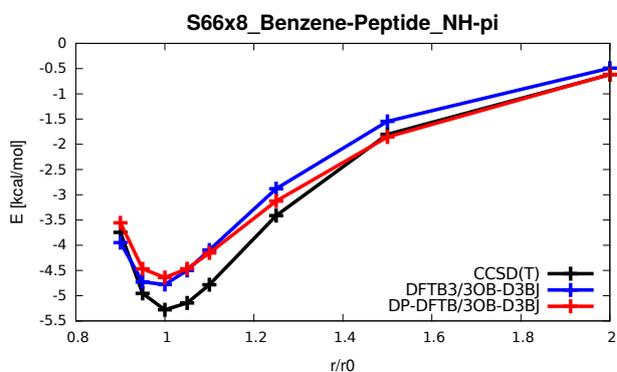
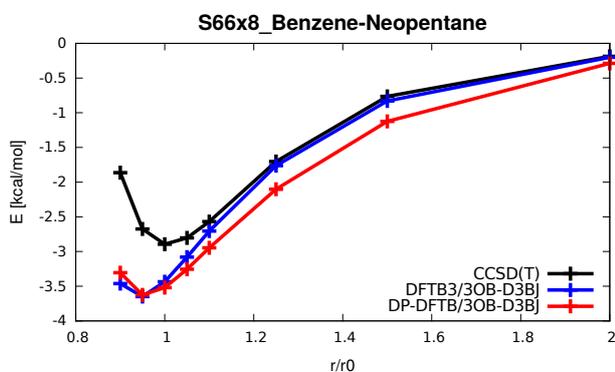
Table B.2: Predefined DP and D3 parameter space with lower and upper boundaries used for PSO fitting.

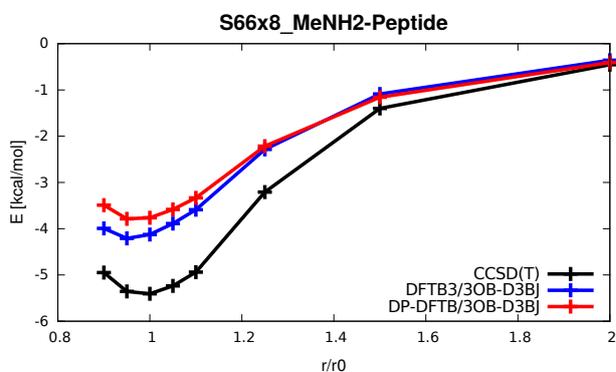
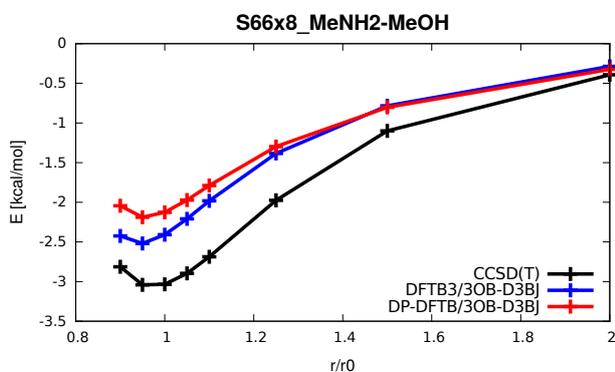
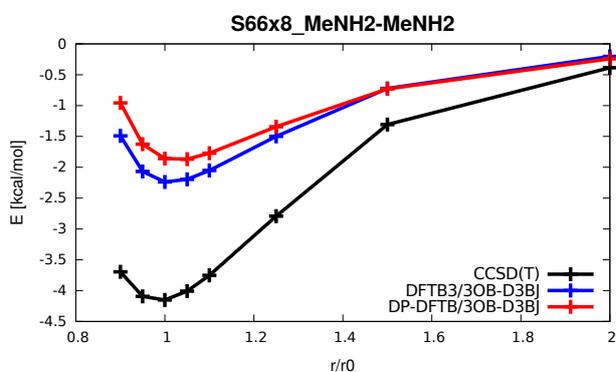
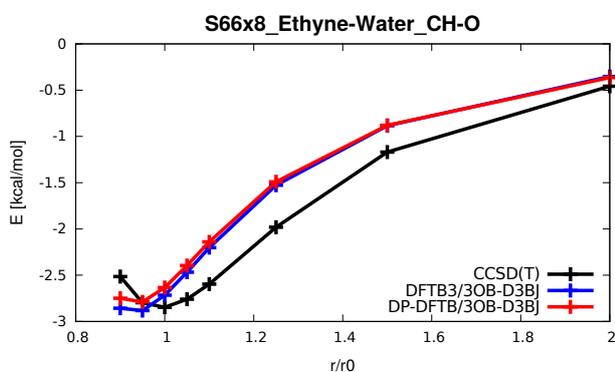
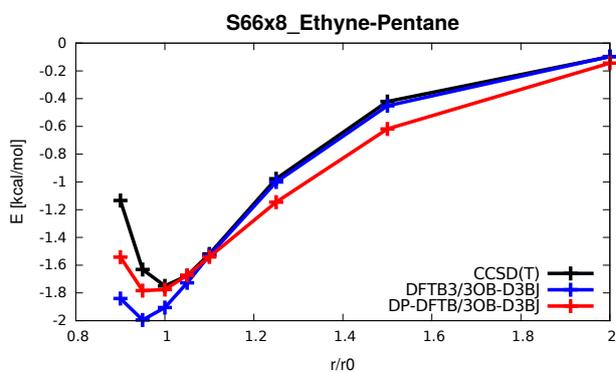
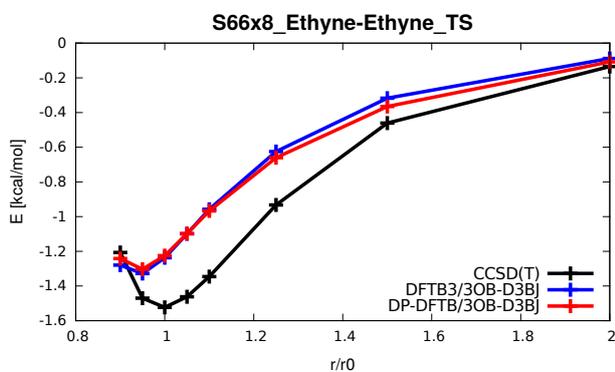
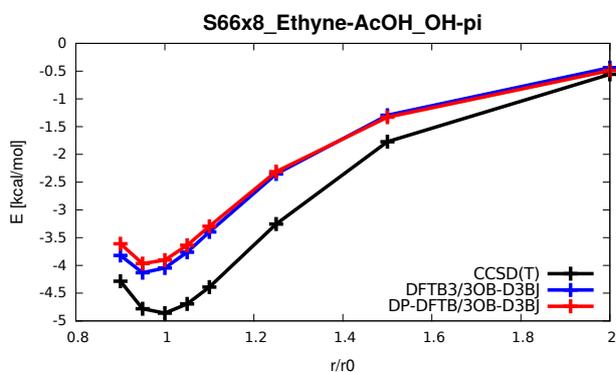
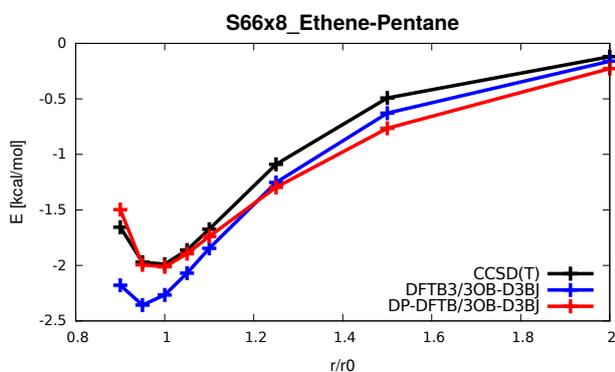
Parameter	min.	max.
ξ	1.0	50.0
d	0.0	5.0
ζ	0.0	10.0
α	0.0	10.0
sC	0.1	50.0
sH	0.1	50.0
sN	0.1	50.0
sO	0.1	50.0
a ₁	0.1	0.9
a ₂	0.1	10.0
s8	0.1	10.0

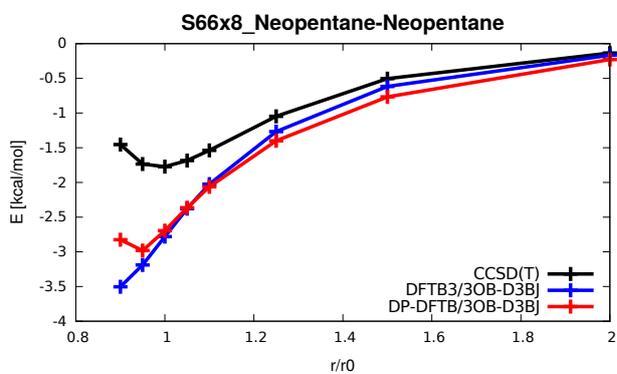
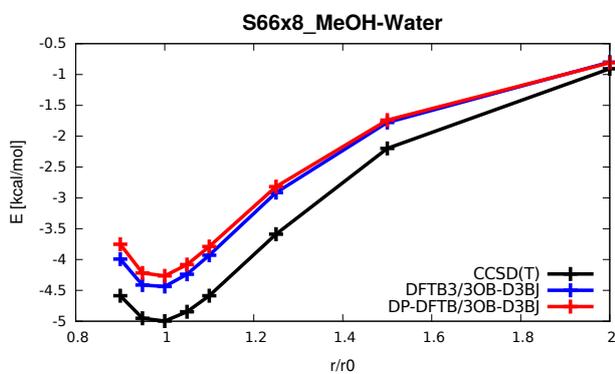
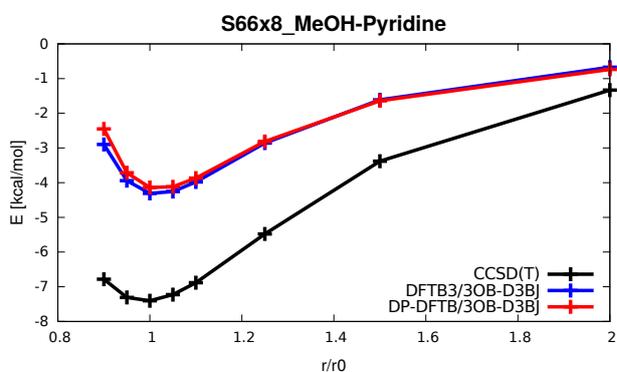
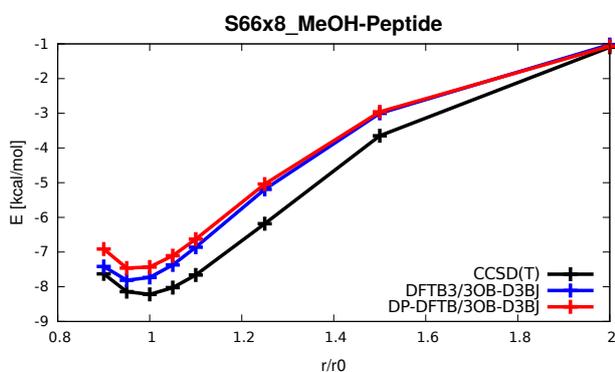
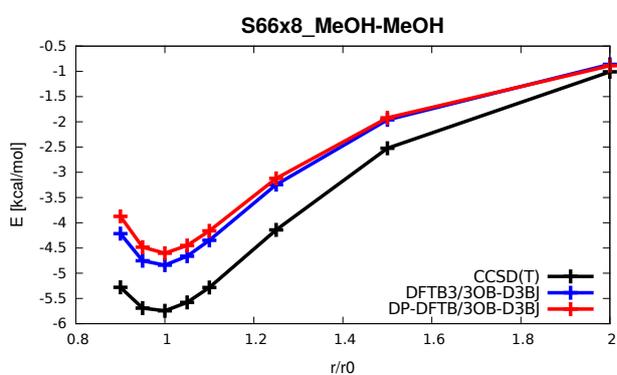
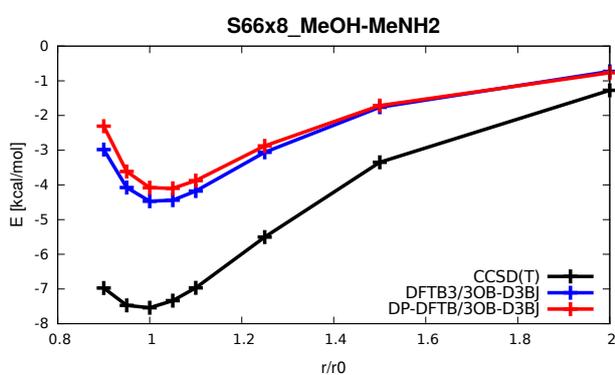
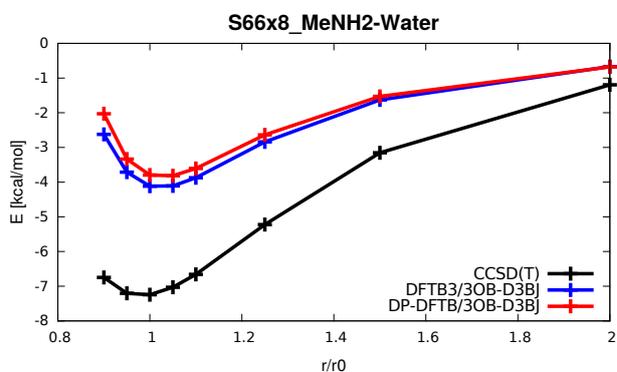
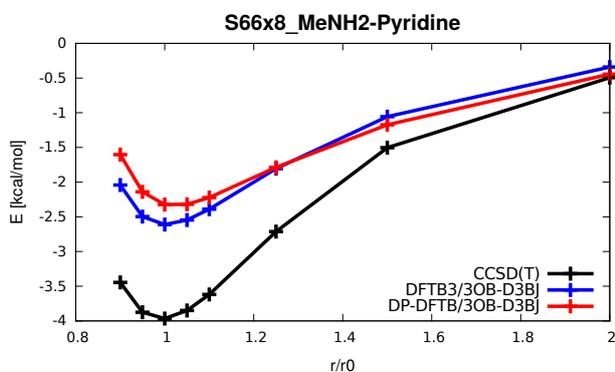


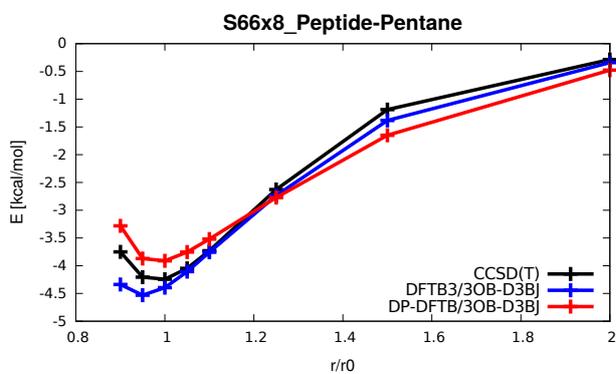
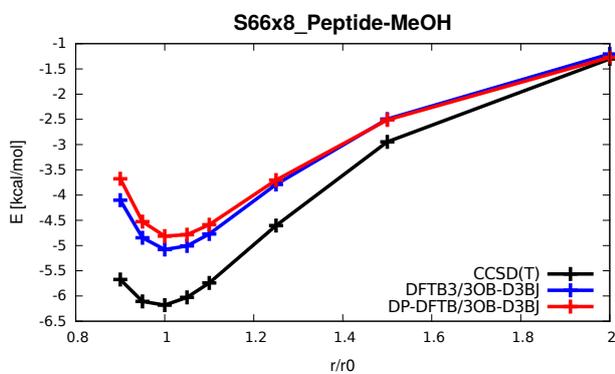
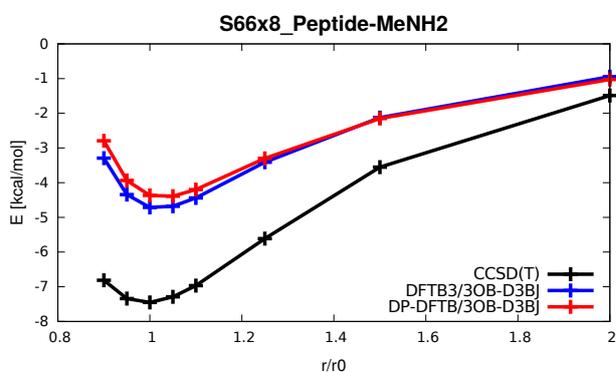
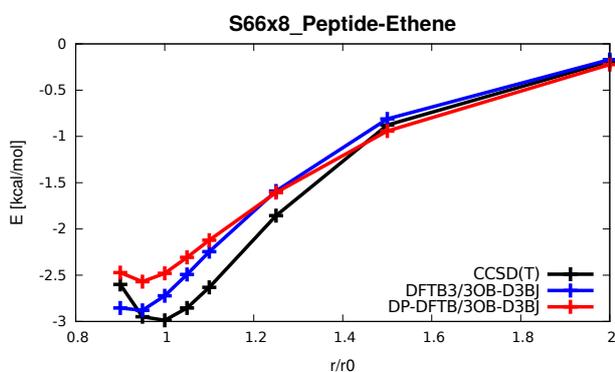
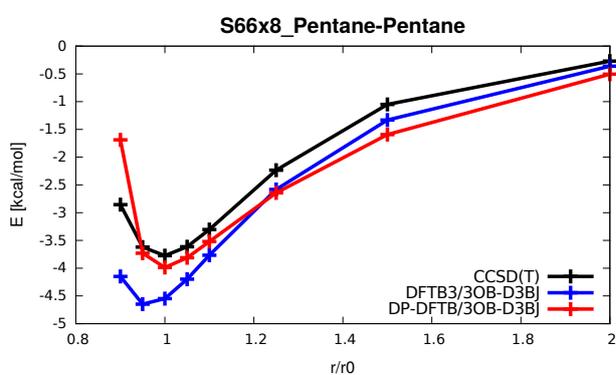
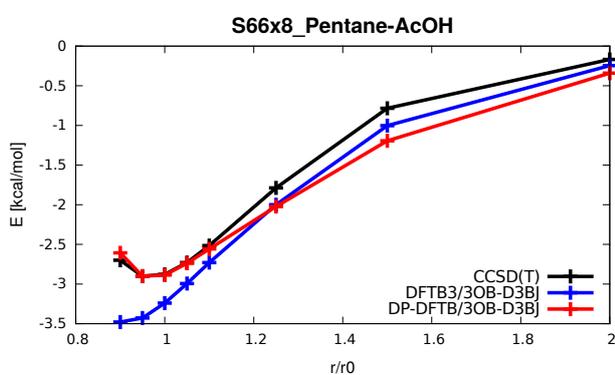
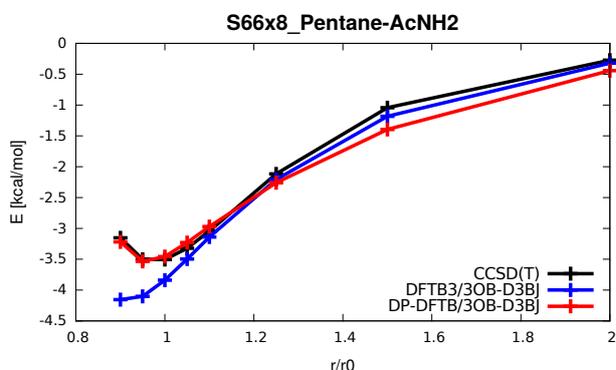
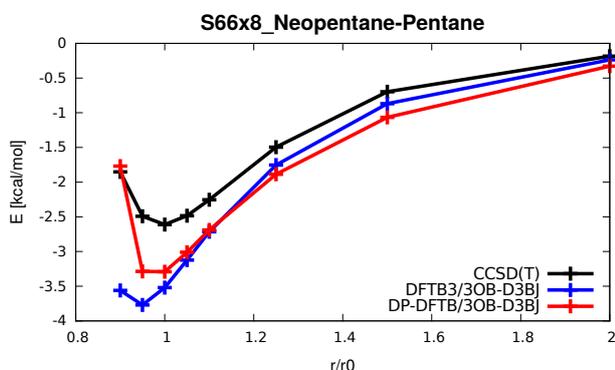


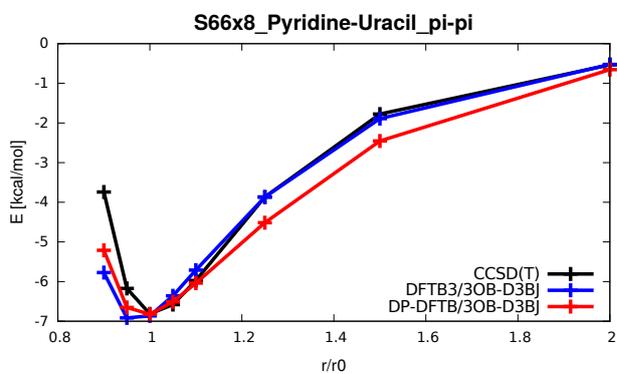
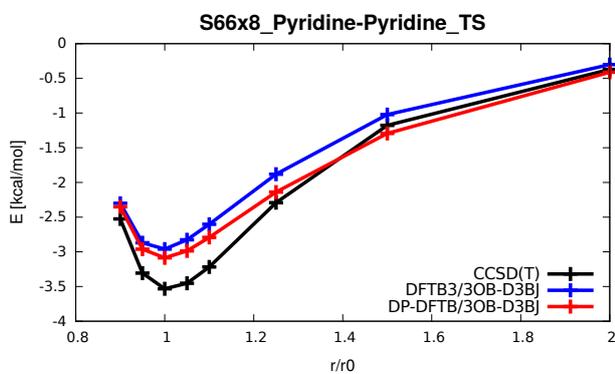
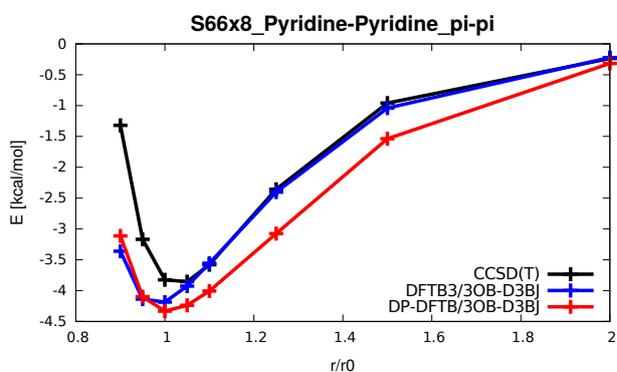
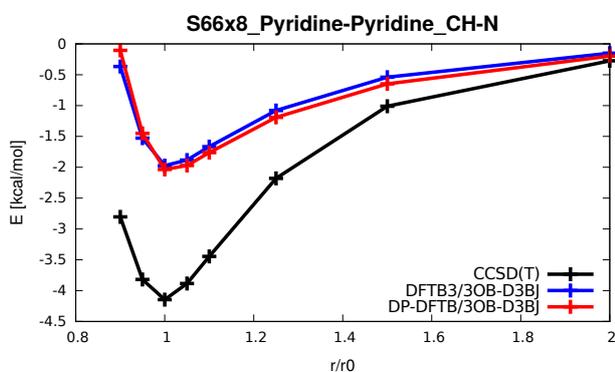
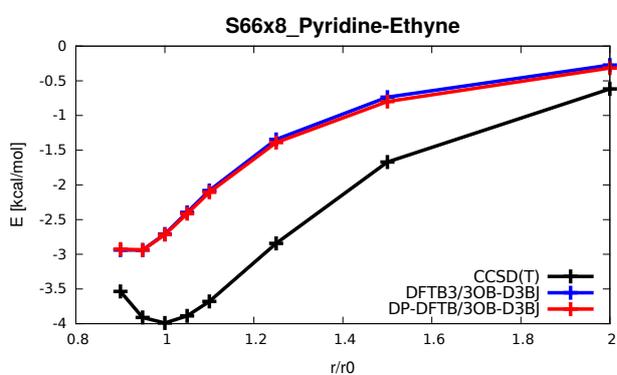
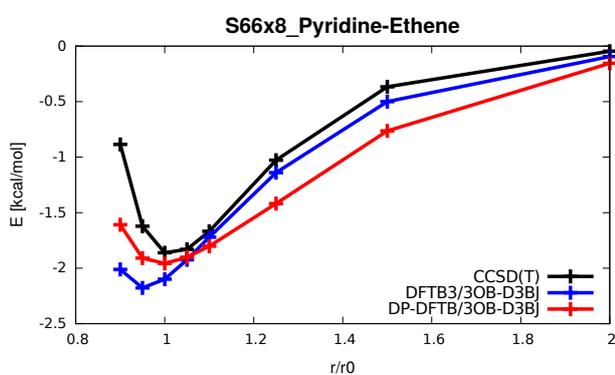
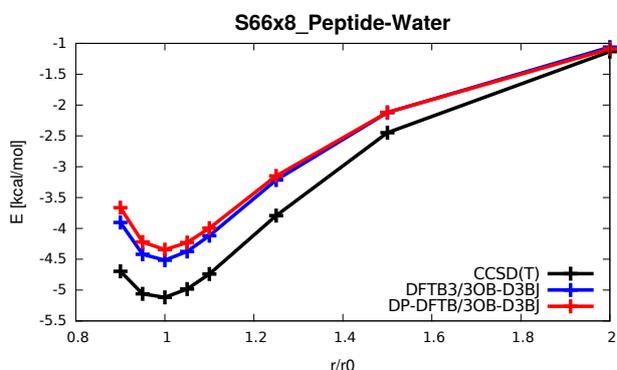
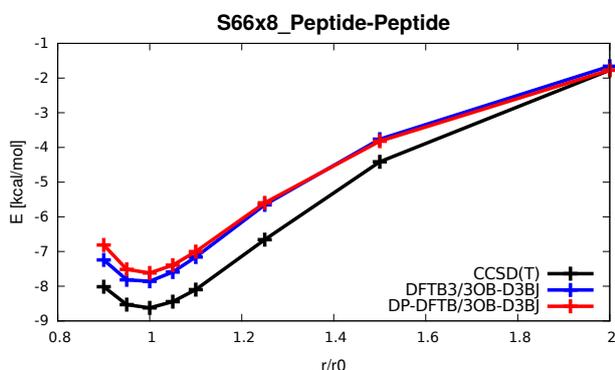


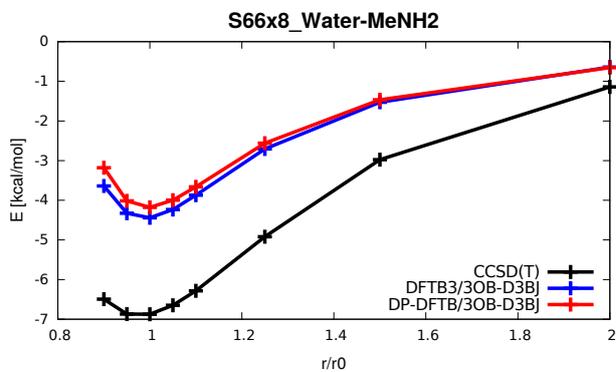
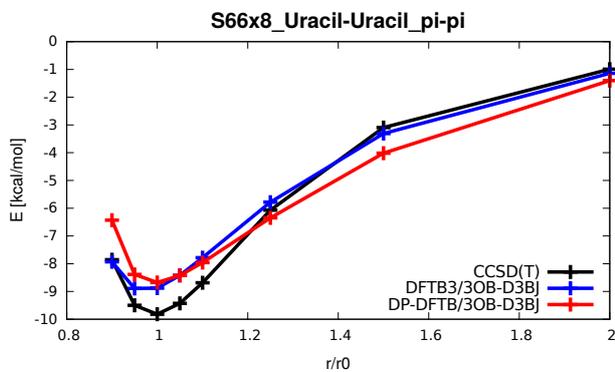
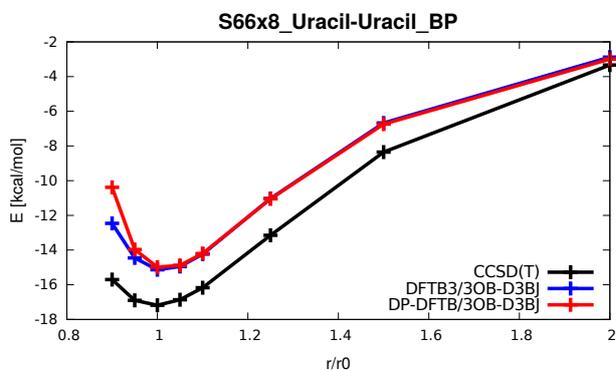
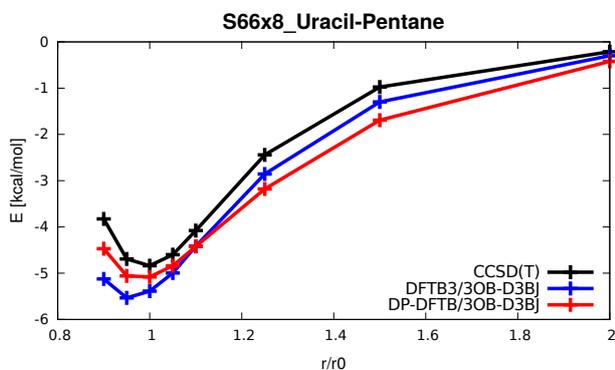
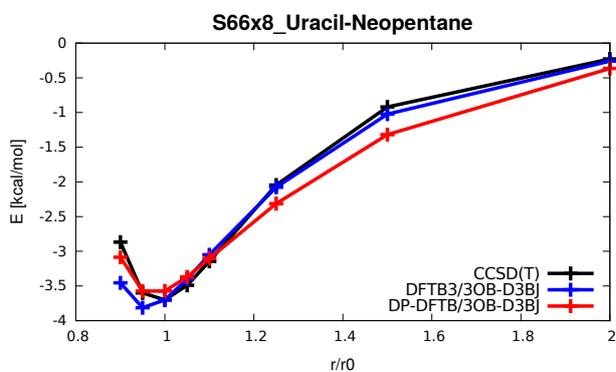
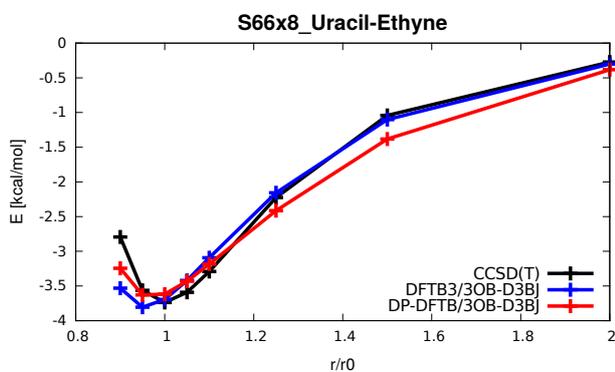
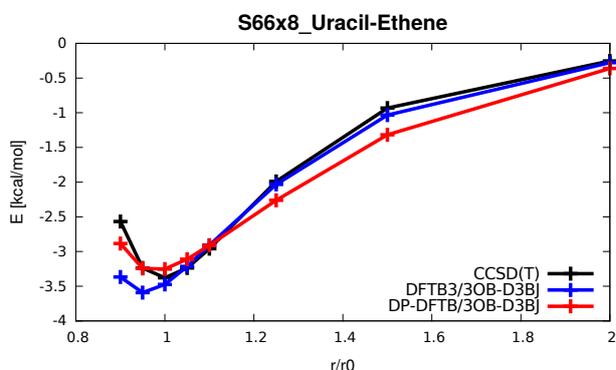
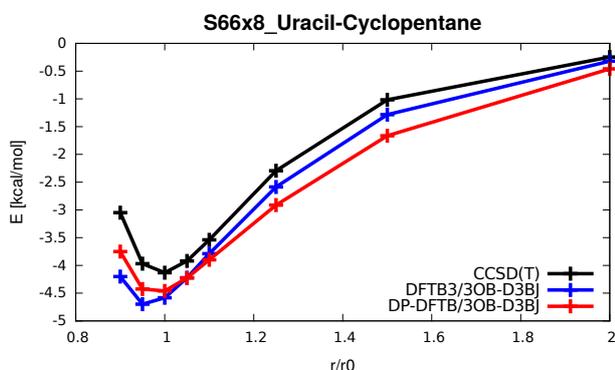


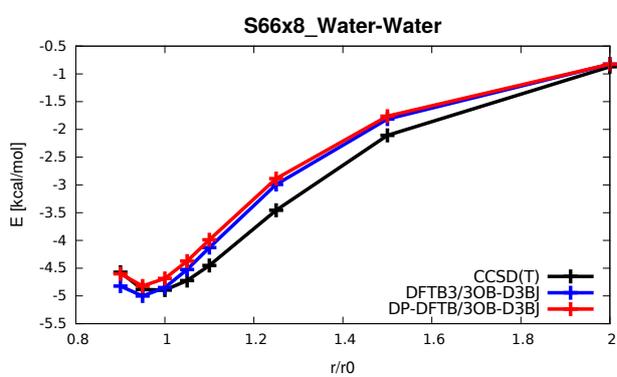
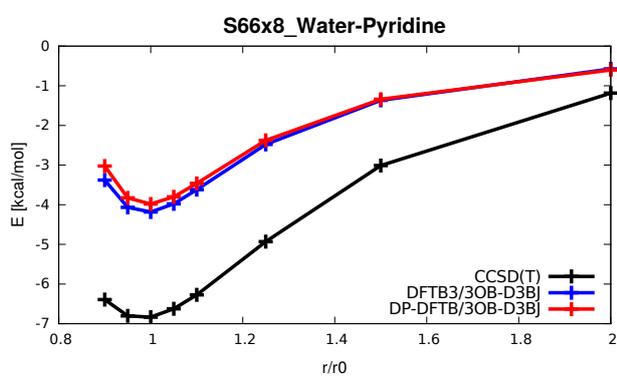
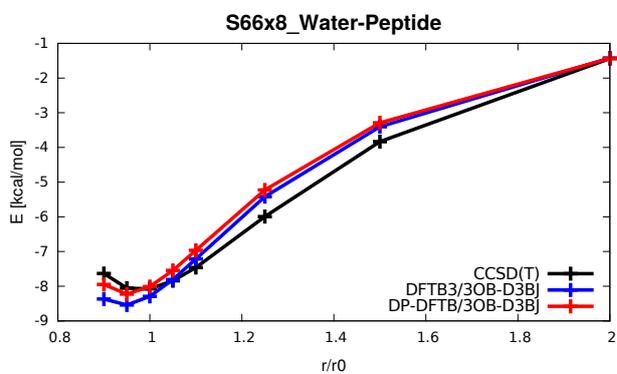
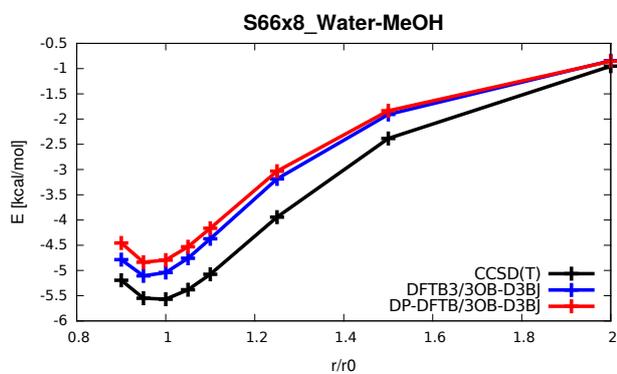












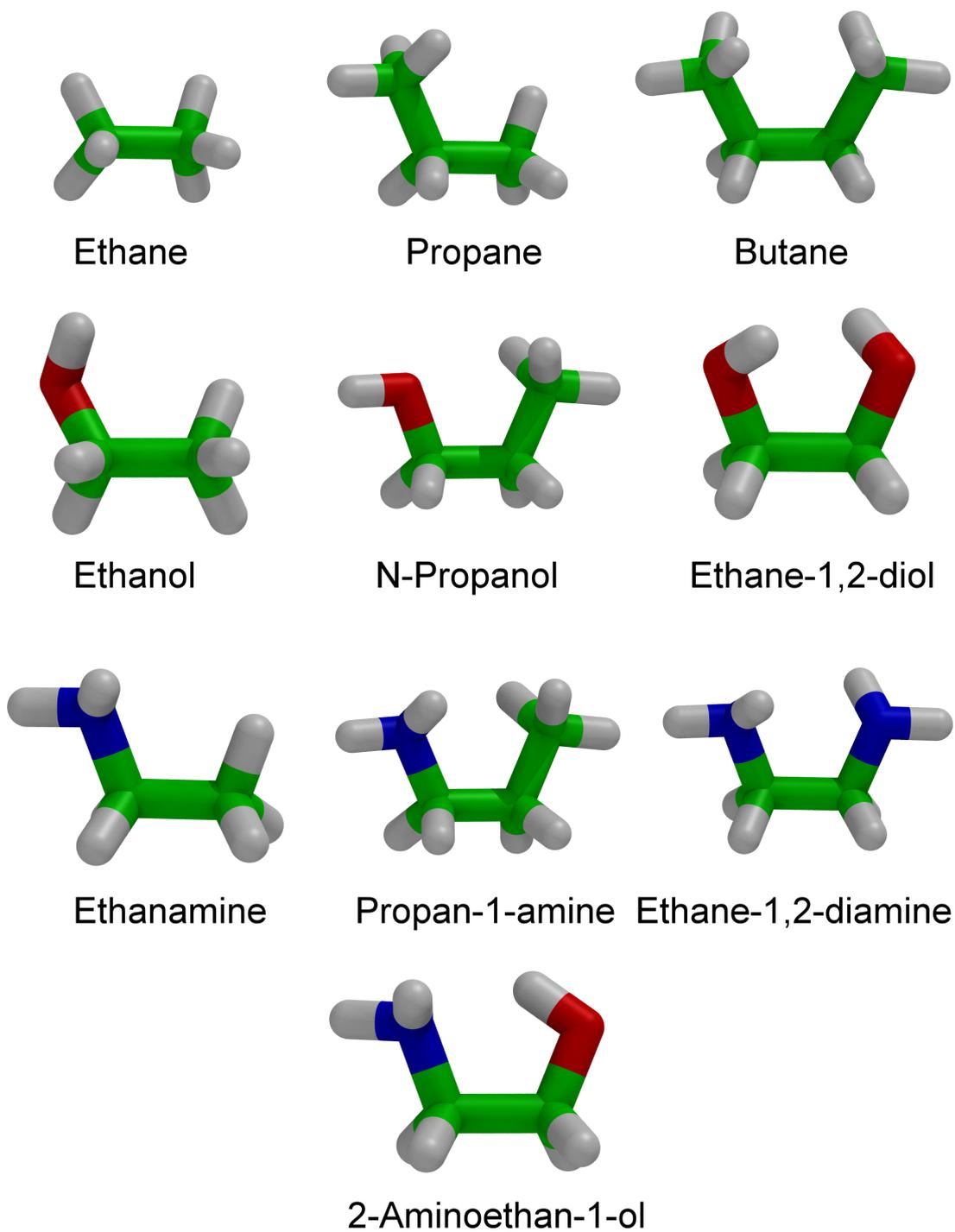


Figure B.1: Composition of the R10 training set.

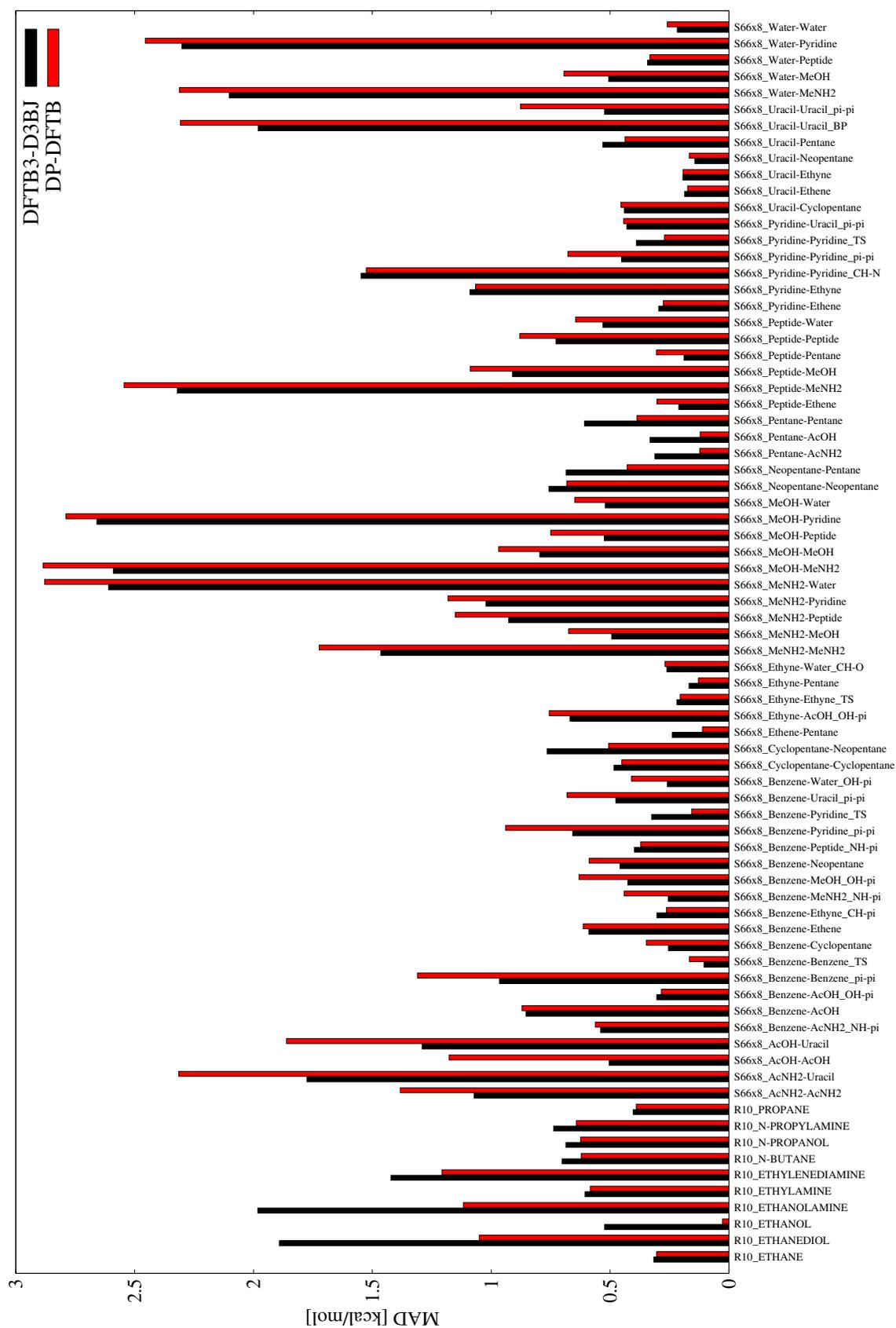


Figure B.3: Mean absolute deviations of DFTB3 (black) and DP-DFTB (red) for the R10 and S66x8 training set systems.

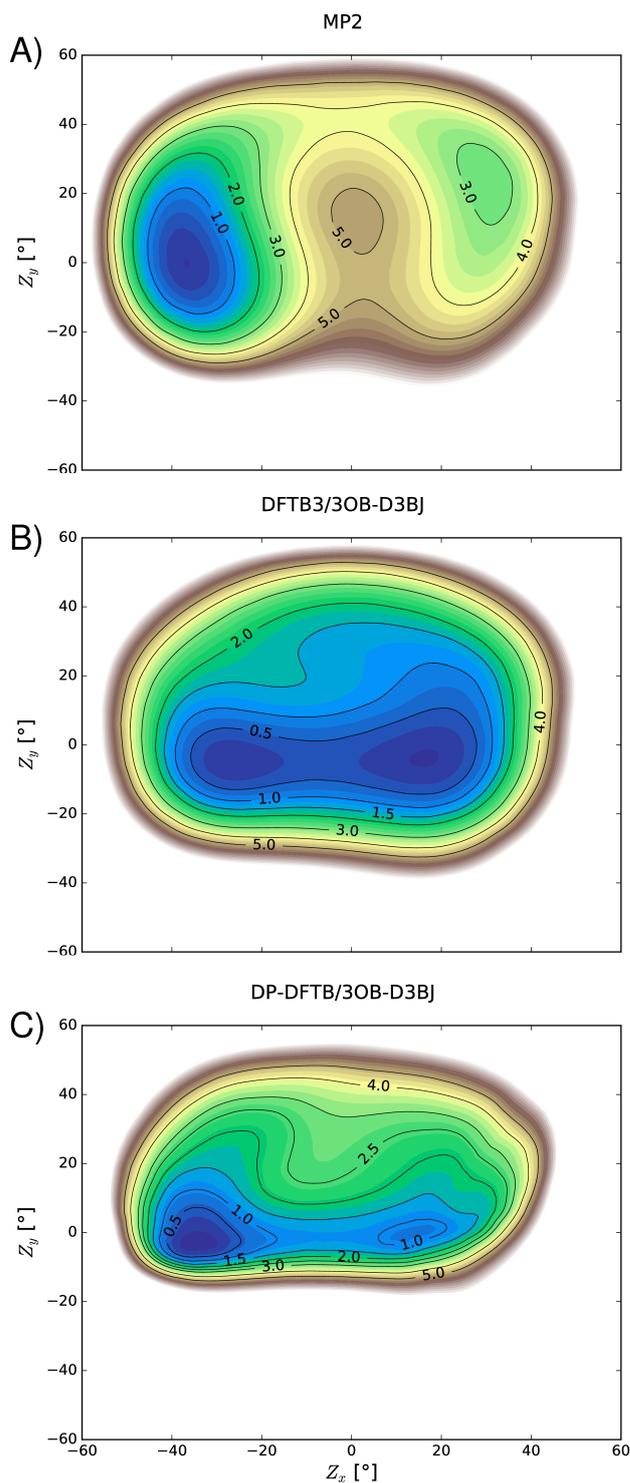


Figure B.4: PES for deoxycytidine (dC) defined by the two puckering parameters Z_x and Z_y . The reference profile calculated with MP2/6-311++G(3df,2p)//MP2/6-31++G(d,p) is shown in A. The corresponding PES calculated with DFTB3/3OB-D3BJ and DP-DFTB/3OB-D3BJ is shown in B and C respectively.

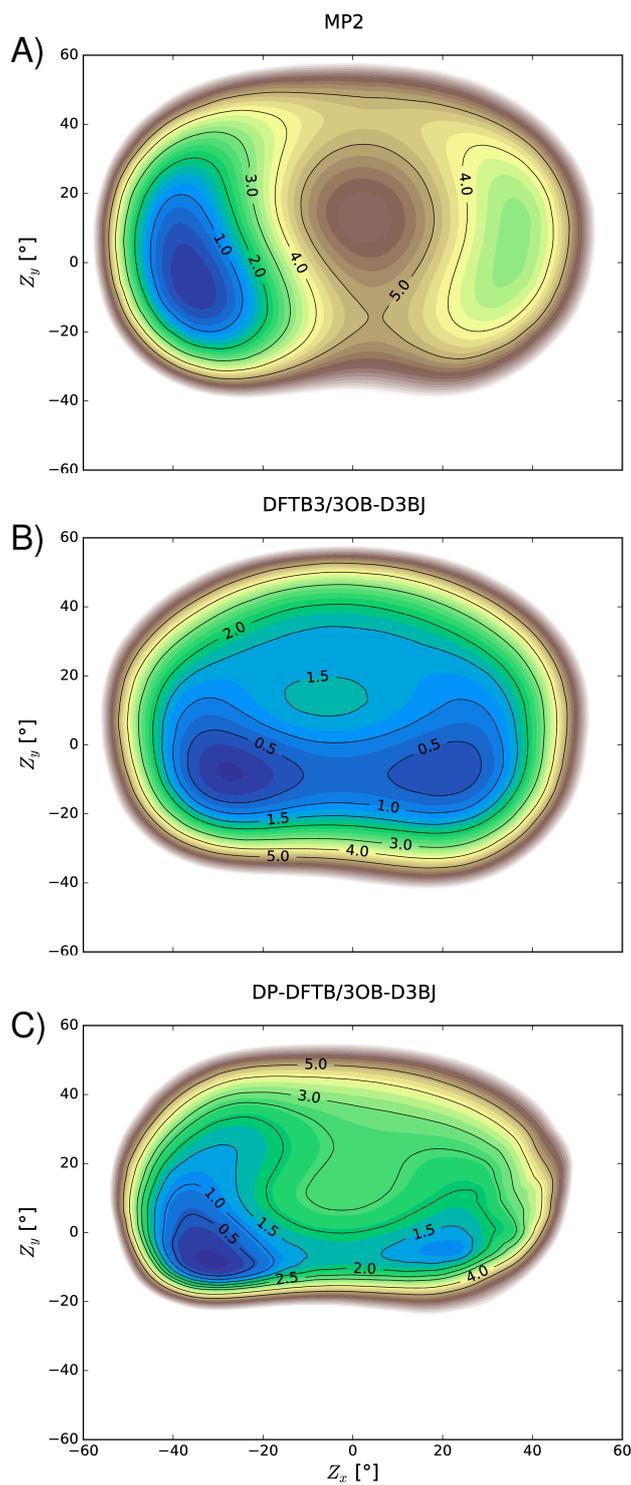


Figure B.5: PES for deoxyguanosine (dG) defined by the two puckering parameters Z_x and Z_y . The reference profile calculated with MP2/6-311++G(3df,2p)//MP2/6-31++G(d,p) is shown in A. The corresponding PES calculated with DFTB3/3OB-D3BJ and DP-DFTB/3OB-D3BJ is shown in B and C respectively.

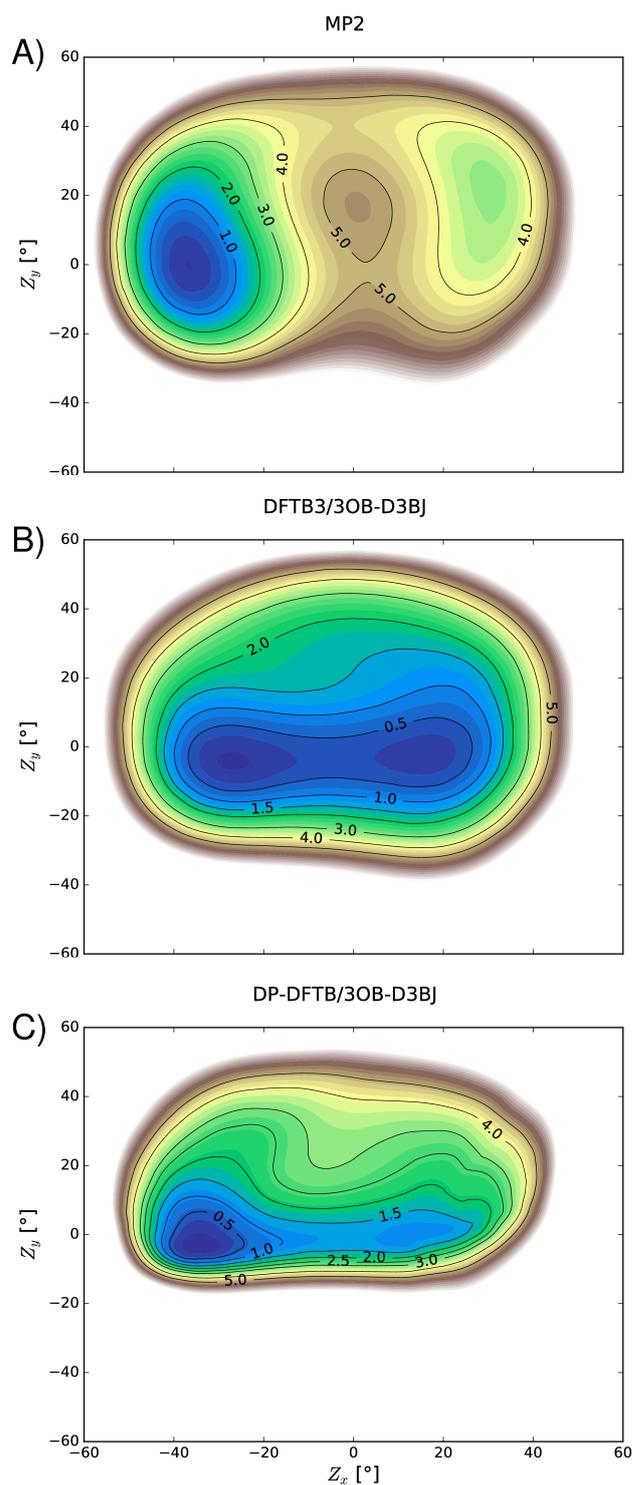


Figure B.6: PES for deoxythymidine (dT) defined by the two puckering parameters Z_x and Z_y . The reference profile calculated with MP2/6-311++G(3df,2p)//MP2/6-31++G(d,p) is shown in A. The corresponding PES calculated with DFTB3/3OB-D3BJ and DP-DFTB/3OB-D3BJ is shown in B and C respectively.

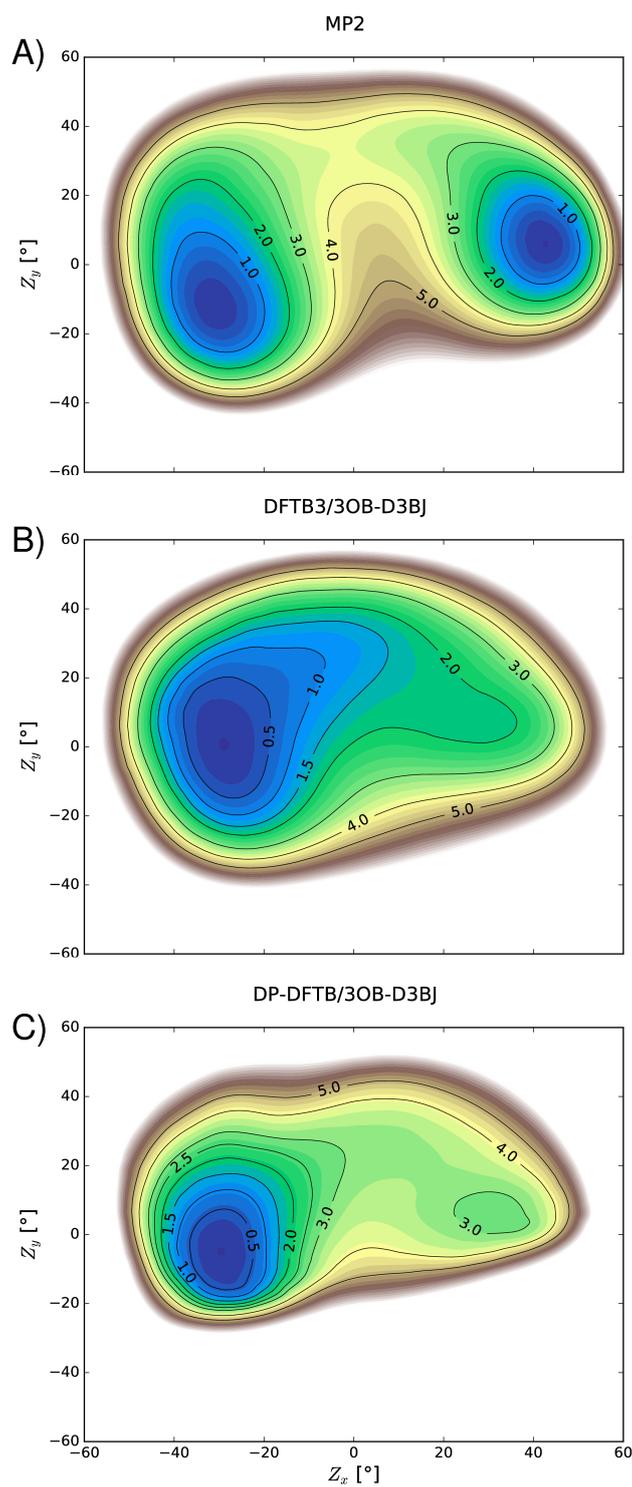


Figure B.7: PES for adenosine (rA) defined by the two pucker parameters Z_x and Z_y . The reference profile calculated with MP2/6-311++G(3df,2p)//MP2/6-31++G(d,p) is shown in A. The corresponding PES calculated with DFTB3/3OB-D3BJ and DP-DFTB/3OB-D3BJ is shown in B and C respectively.

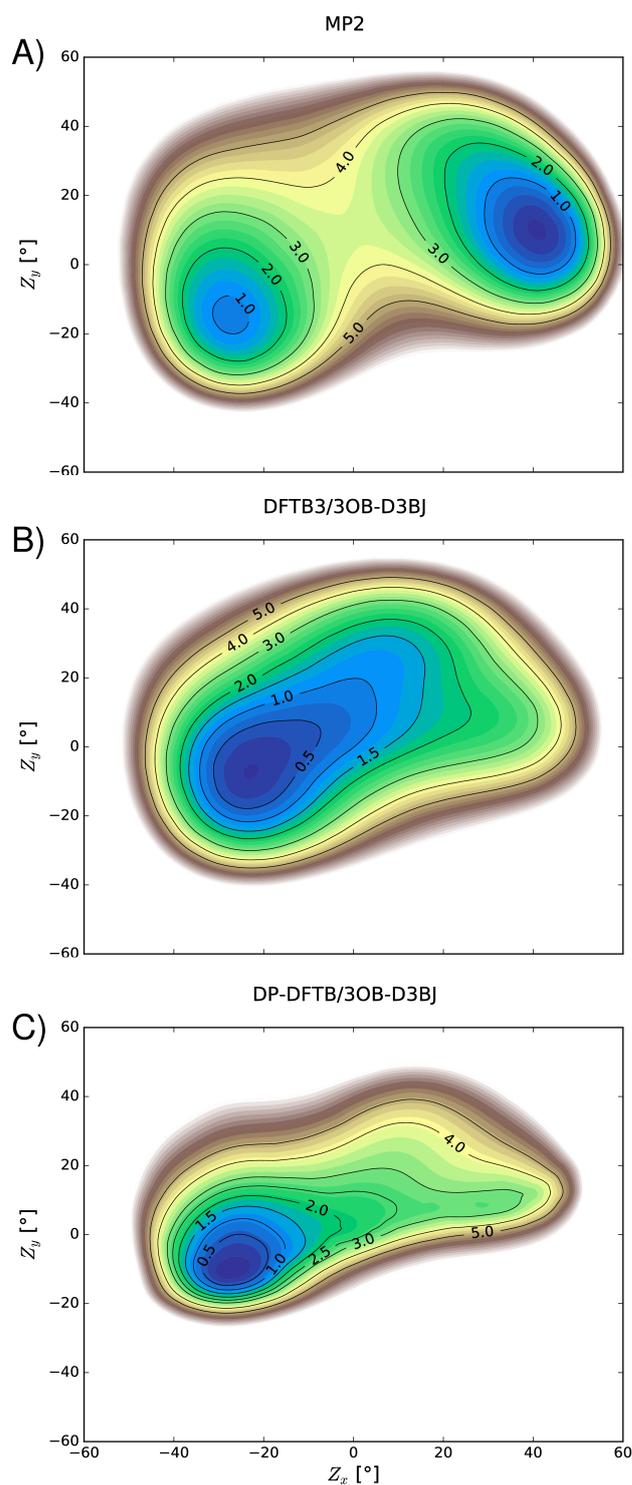


Figure B.8: PES for cytidine (rC) defined by the two puckering parameters Z_x and Z_y . The reference profile calculated with MP2/6-311++G(3df,2p)//MP2/6-31++G(d,p) is shown in A. The corresponding PES calculated with DFTB3/3OB-D3BJ and DP-DFTB/3OB-D3BJ is shown in B and C respectively.

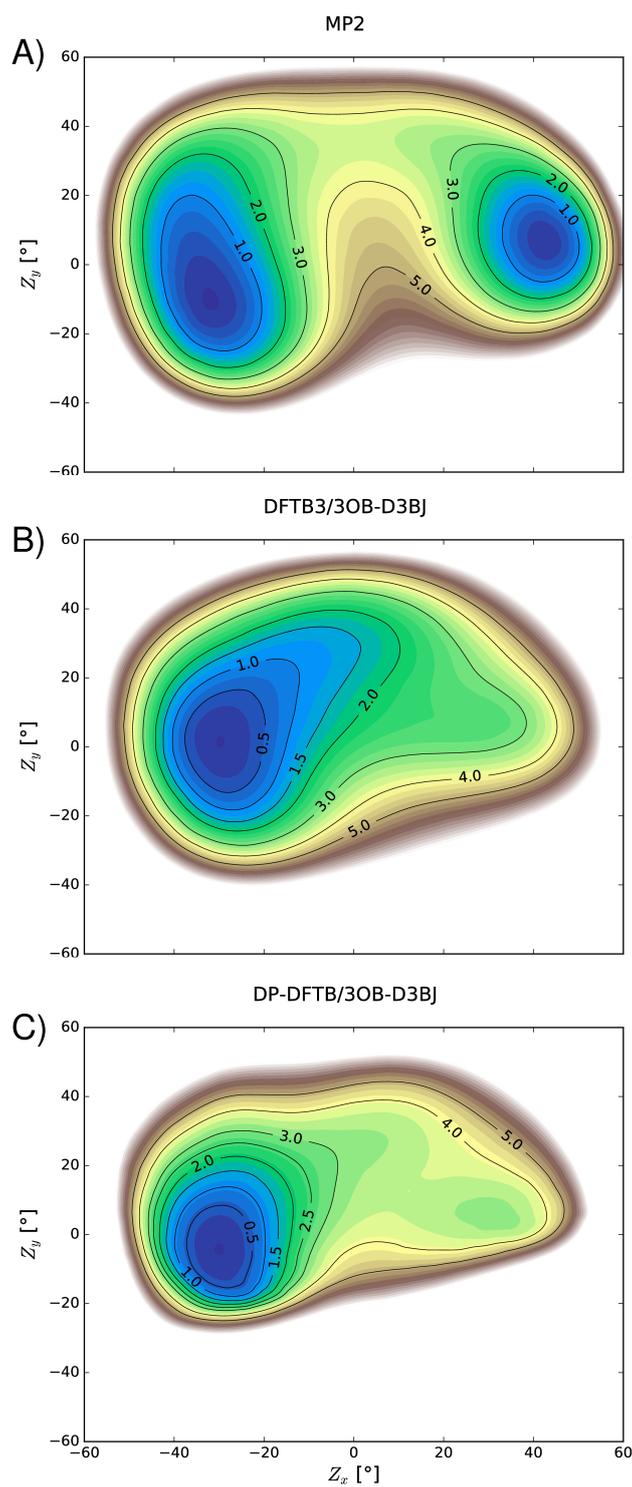


Figure B.9: PES for guanosine (rG) defined by the two pucker parameters Z_x and Z_y . The reference profile calculated with MP2/6-311++G(3df,2p)//MP2/6-31++G(d,p) is shown in A. The corresponding PES calculated with DFTB3/3OB-D3BJ and DP-DFTB/3OB-D3BJ is shown in B and C respectively.

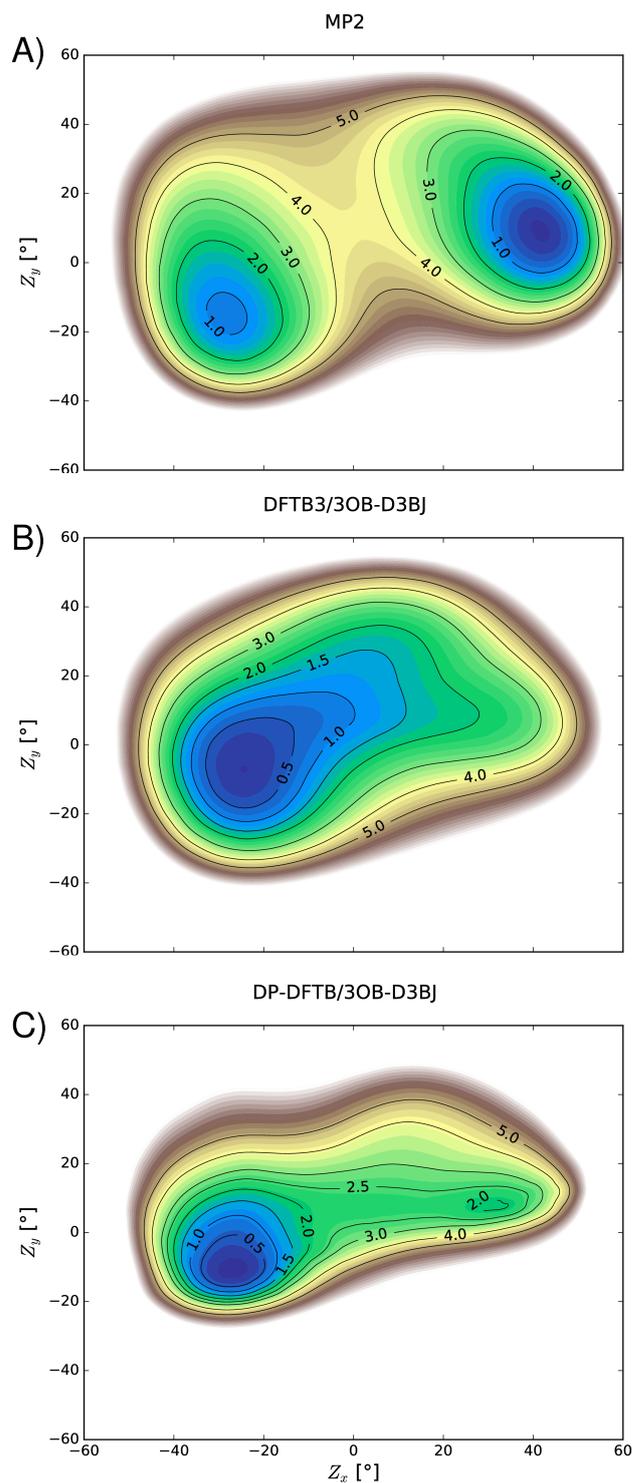
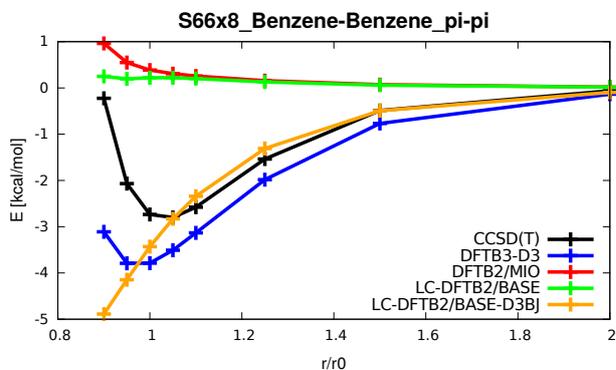
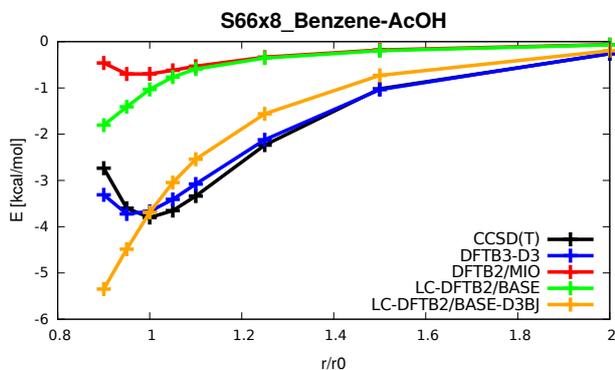
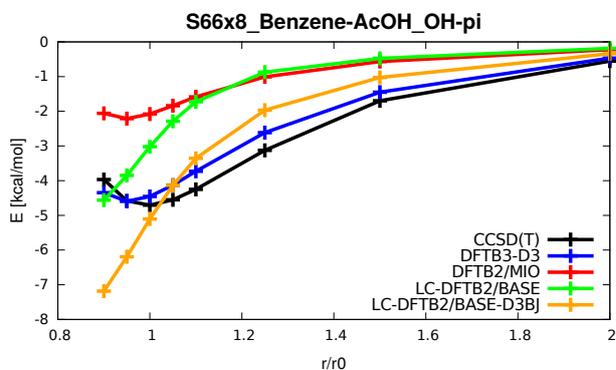
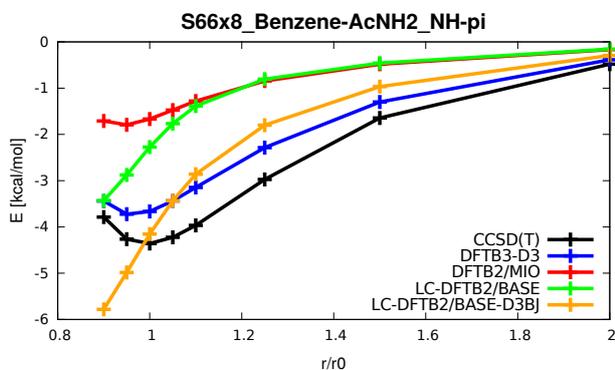
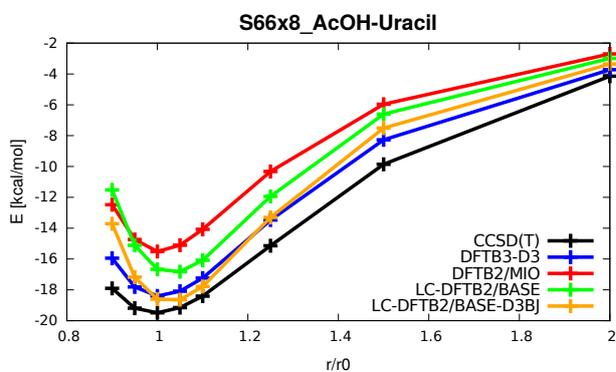
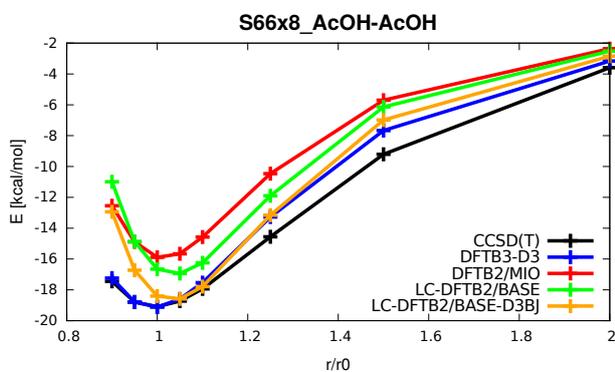
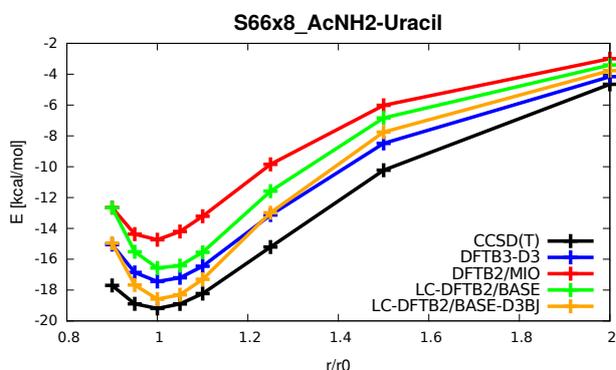
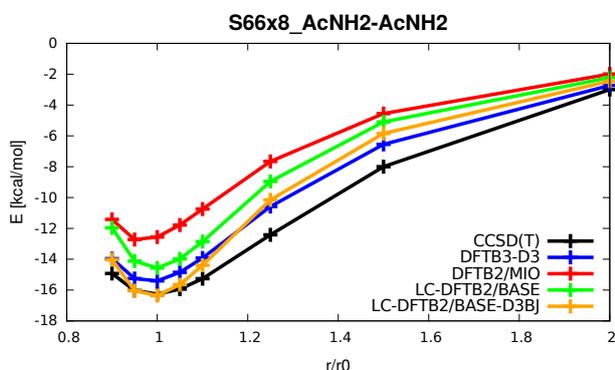
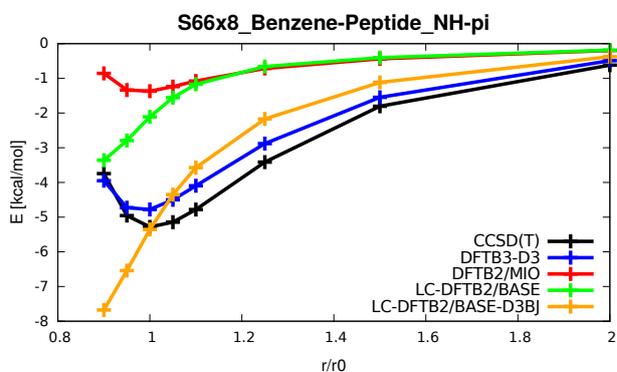
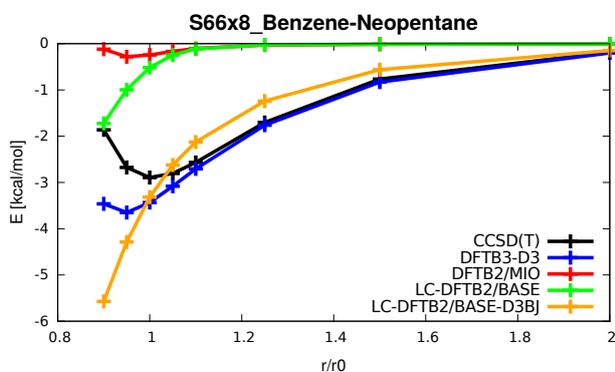
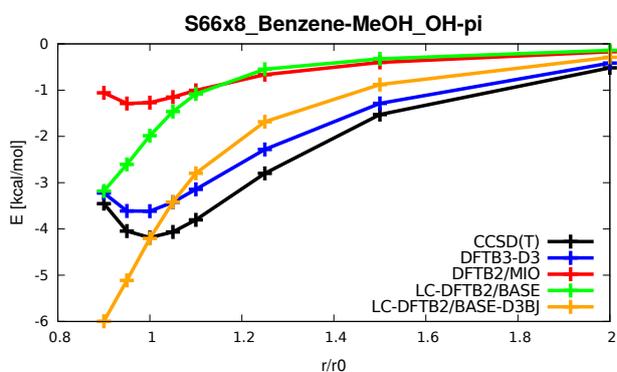
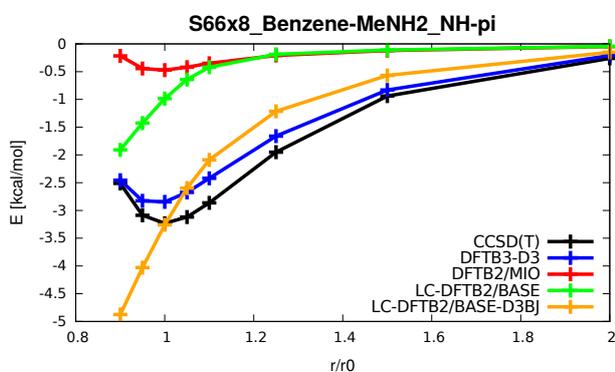
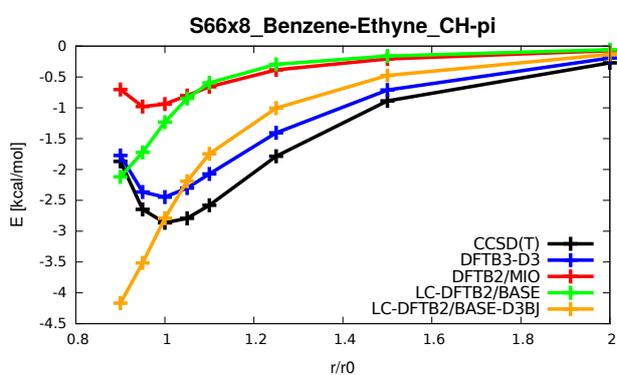
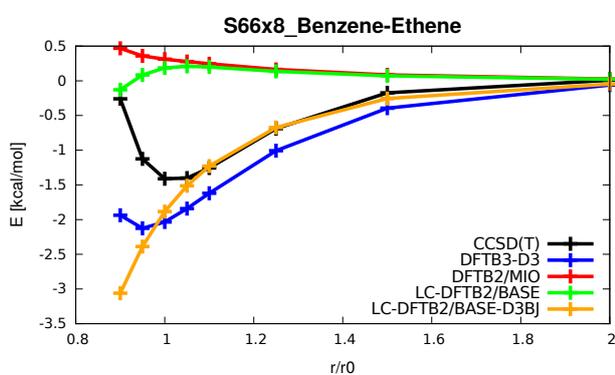
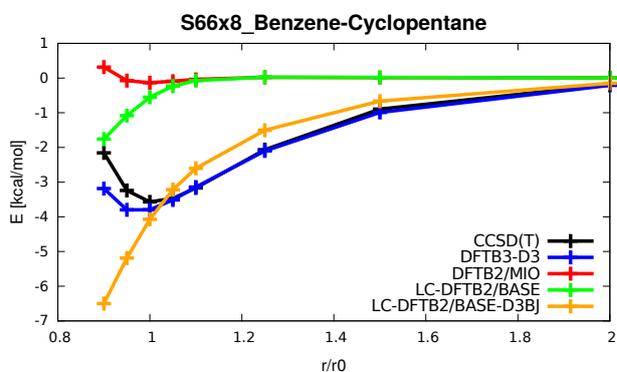
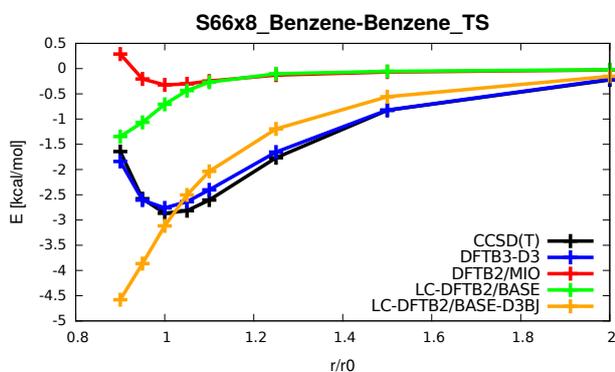


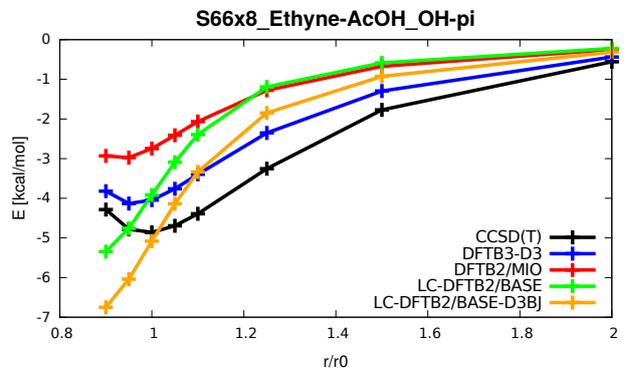
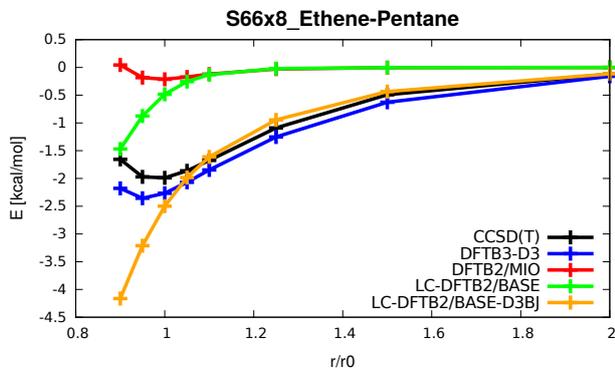
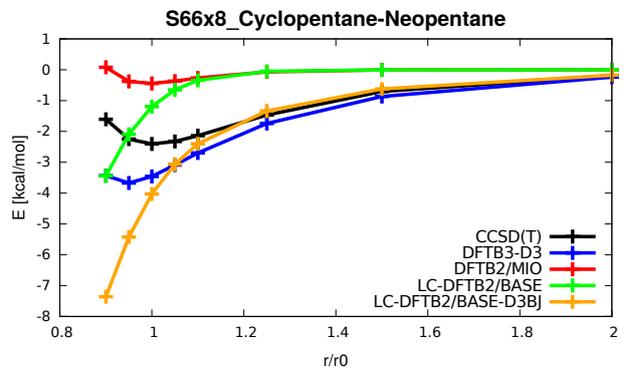
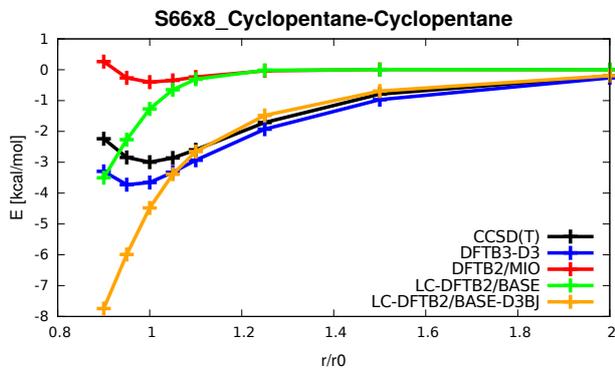
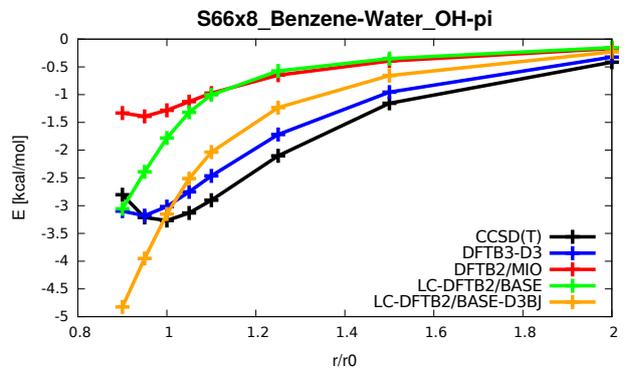
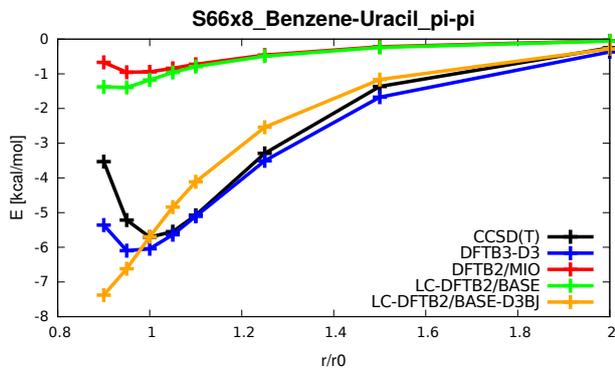
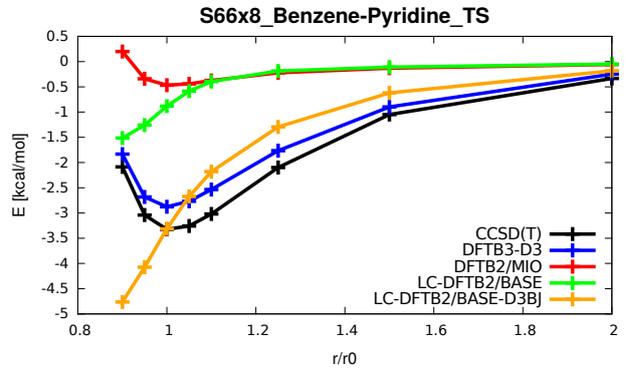
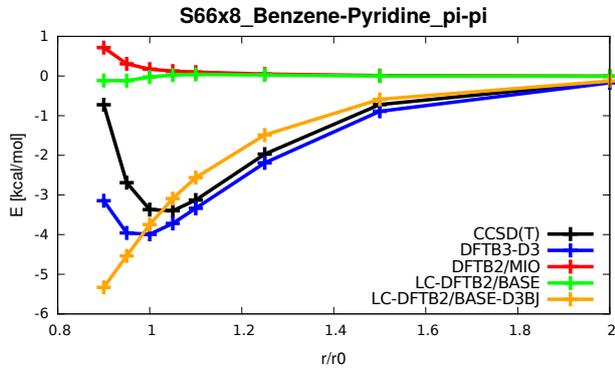
Figure B.10: PES for uridine (rU) defined by the two puckering parameters Z_x and Z_y . The reference profile calculated with MP2/6-311++G(3df,2p)//MP2/6-31++G(d,p) is shown in A. The corresponding PES calculated with DFTB3/3OB-D3BJ and DP-DFTB/3OB-D3BJ is shown in B and C respectively.

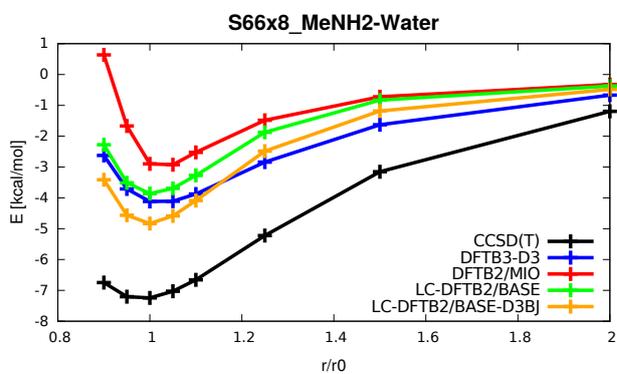
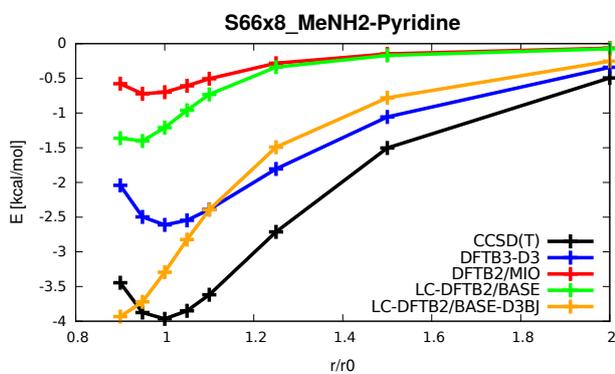
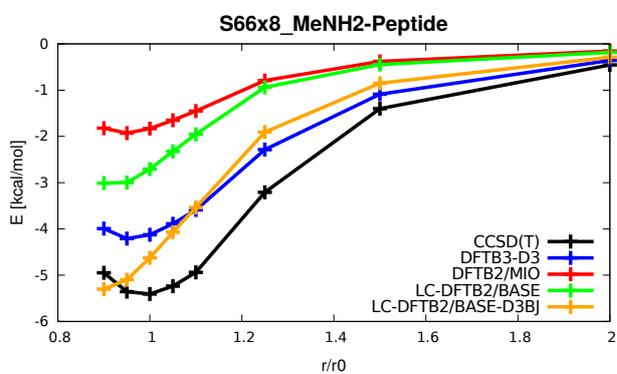
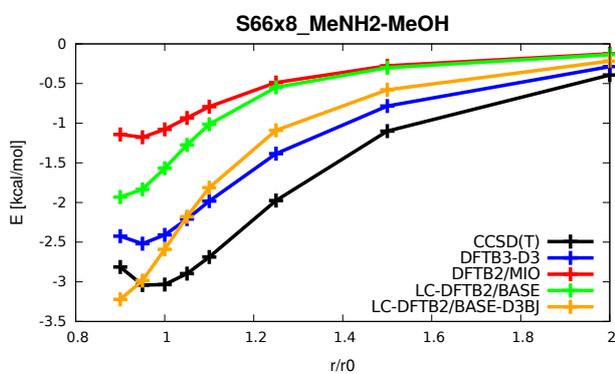
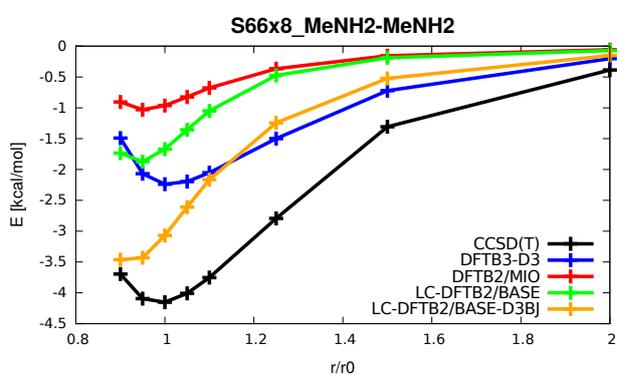
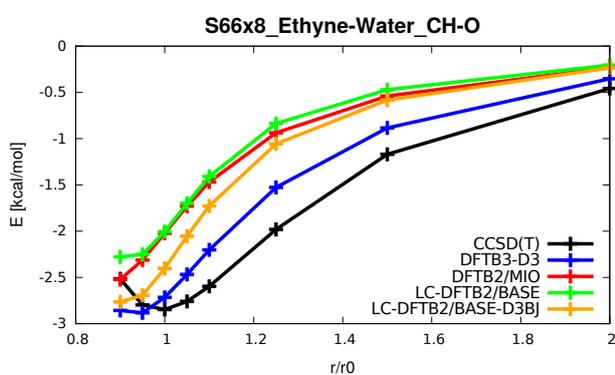
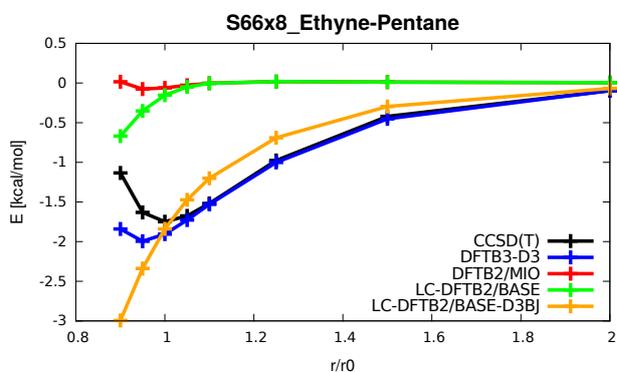
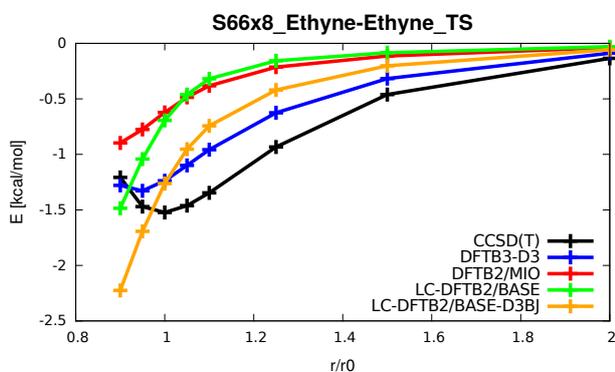
APPENDIX **C** 

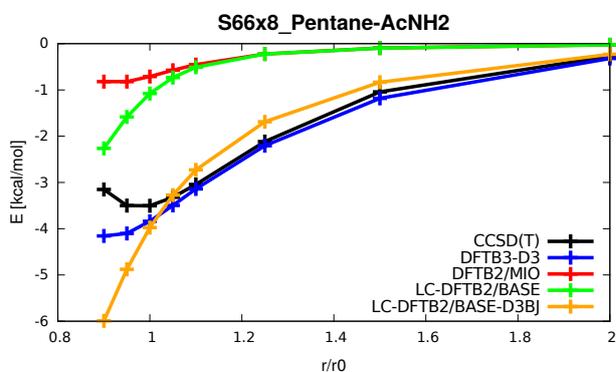
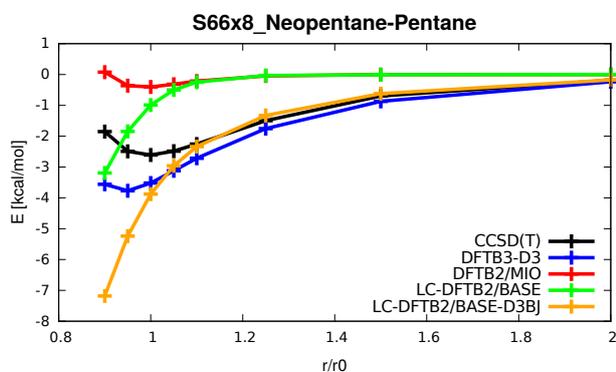
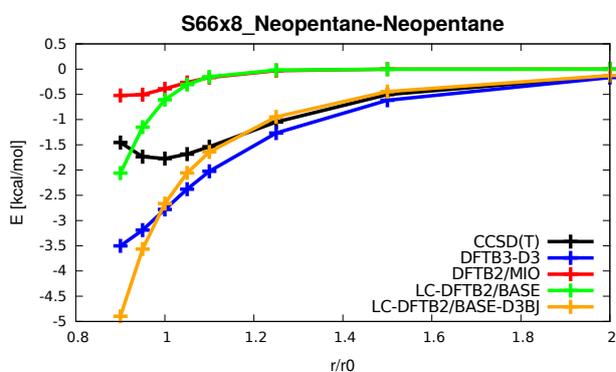
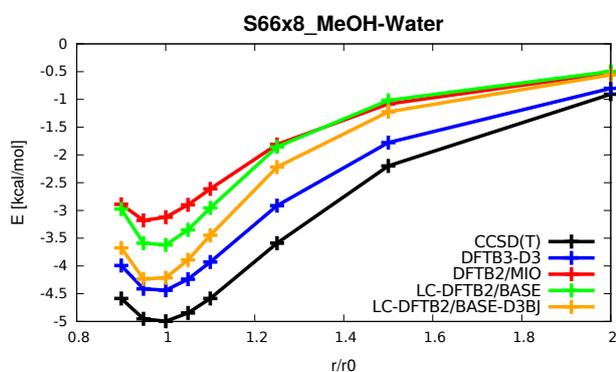
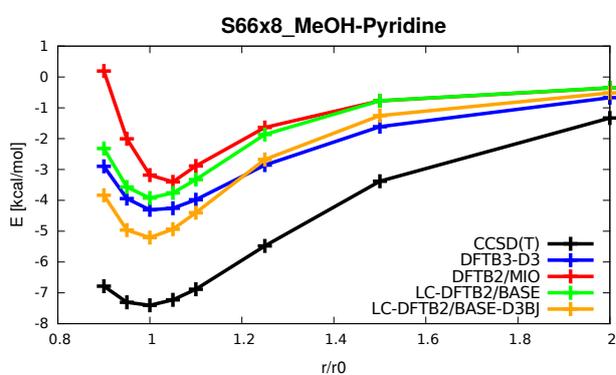
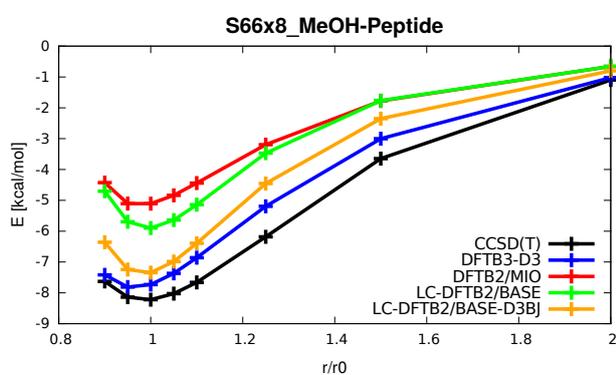
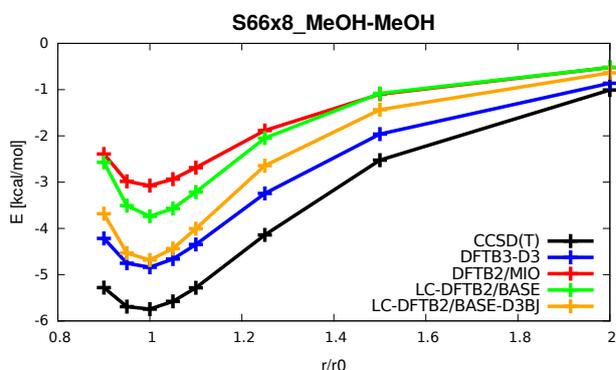
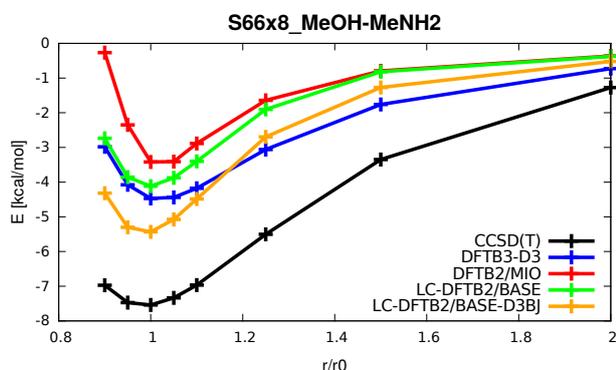
LC-DFTB2

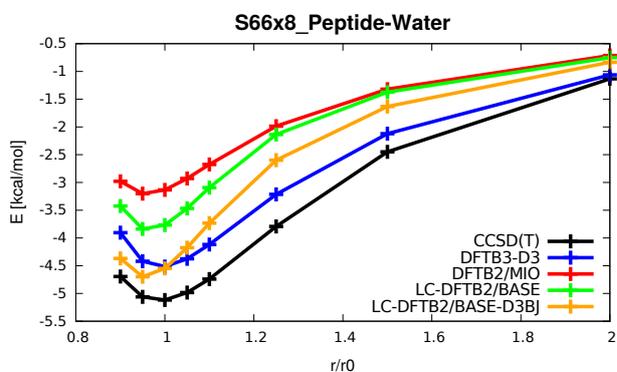
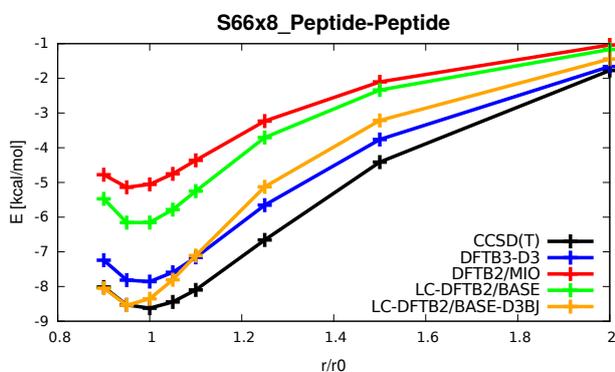
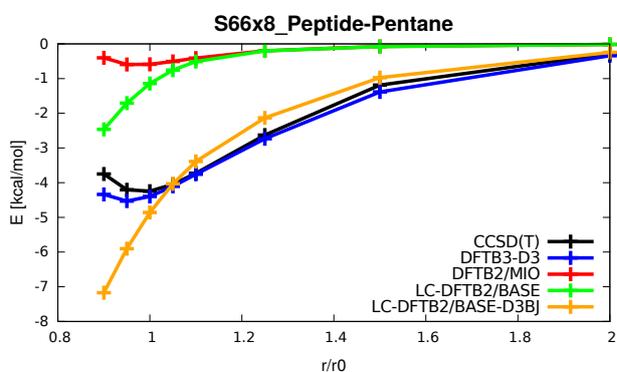
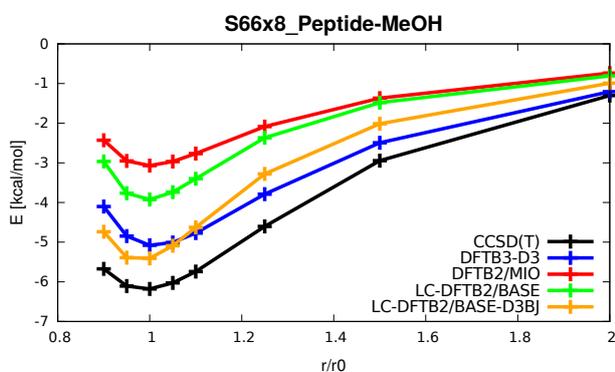
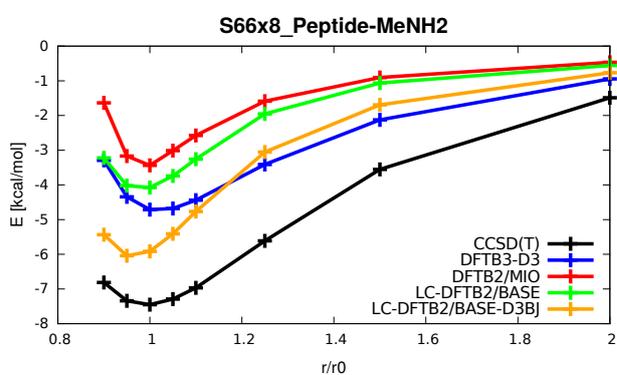
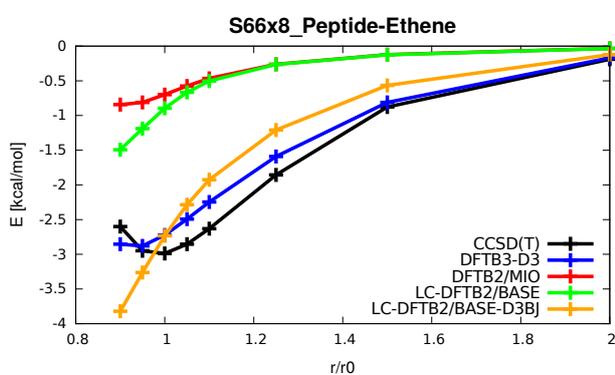
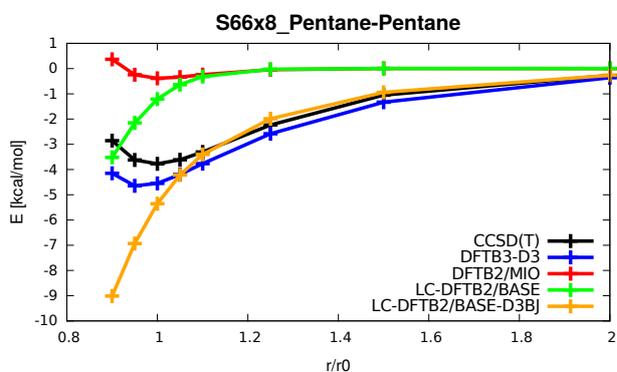
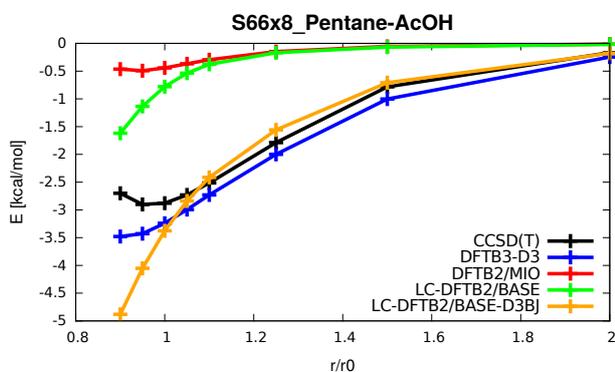


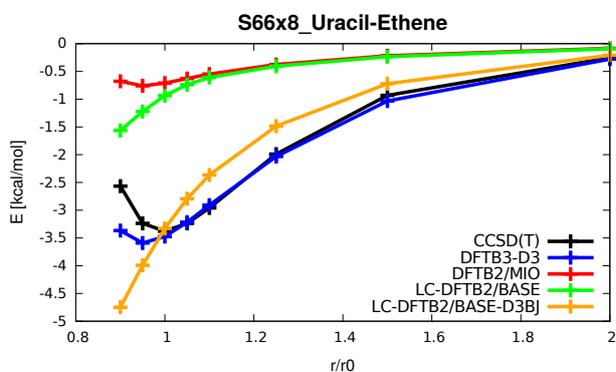
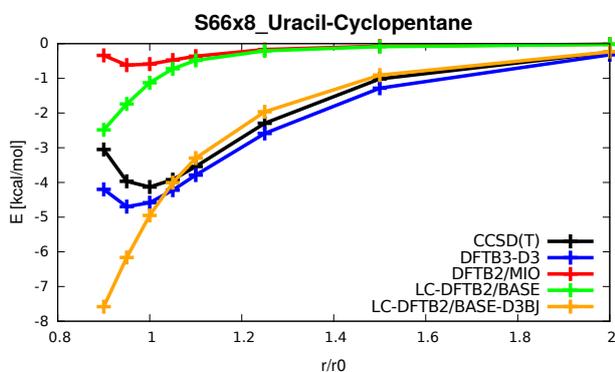
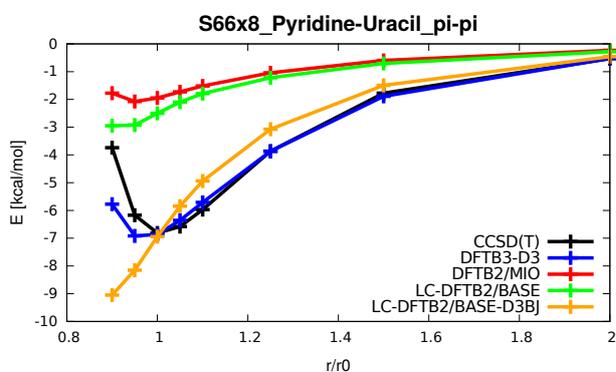
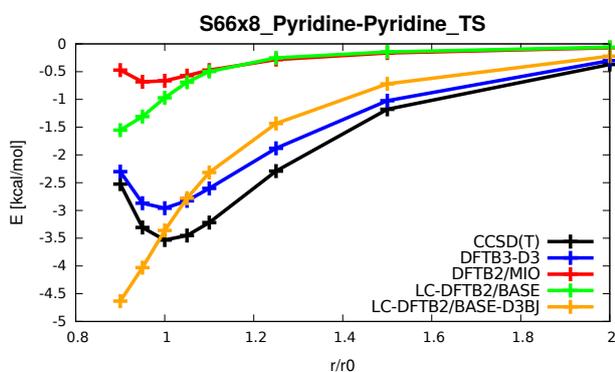
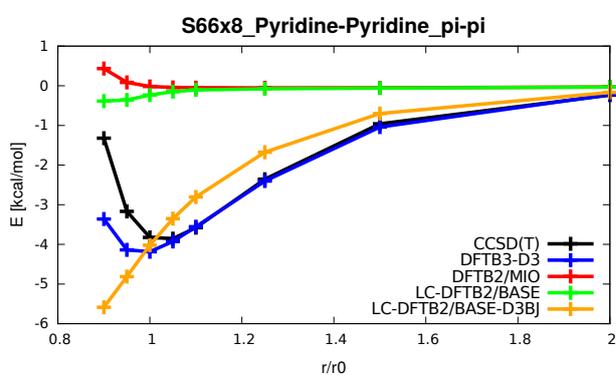
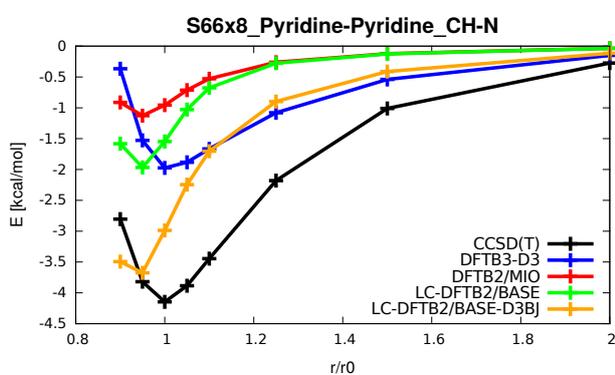
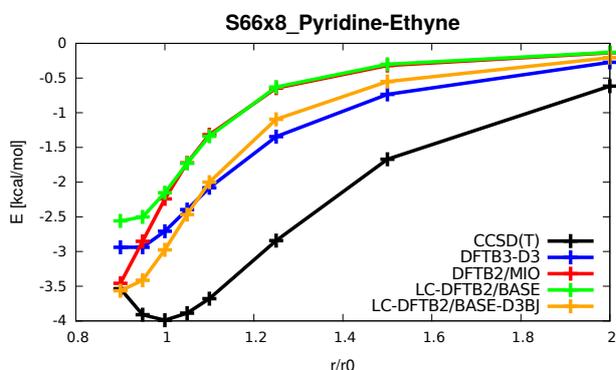
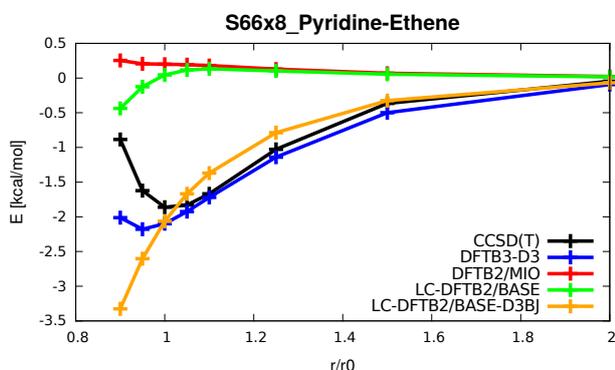


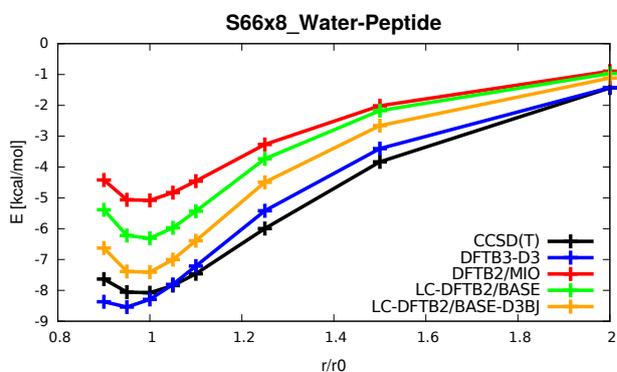
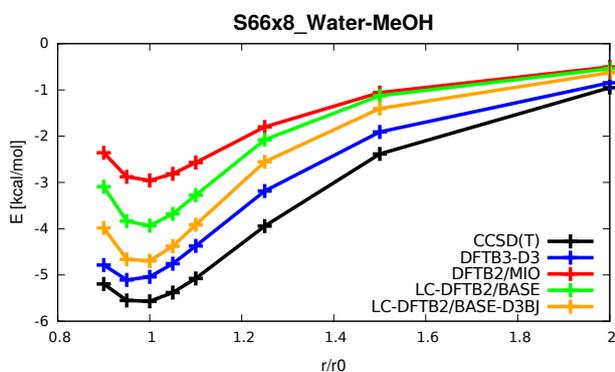
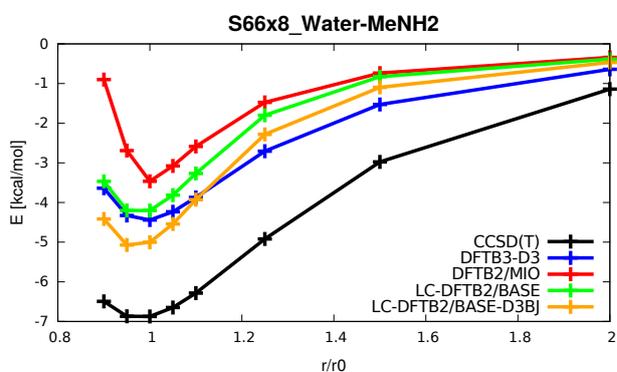
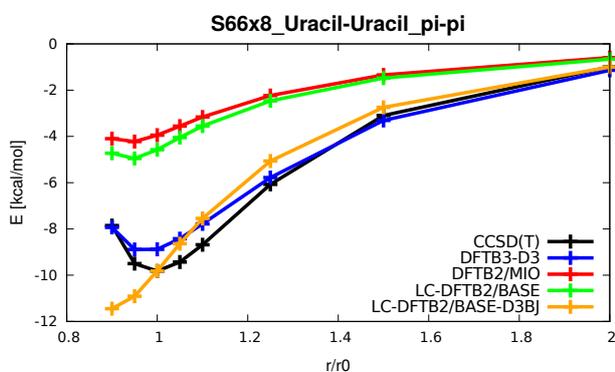
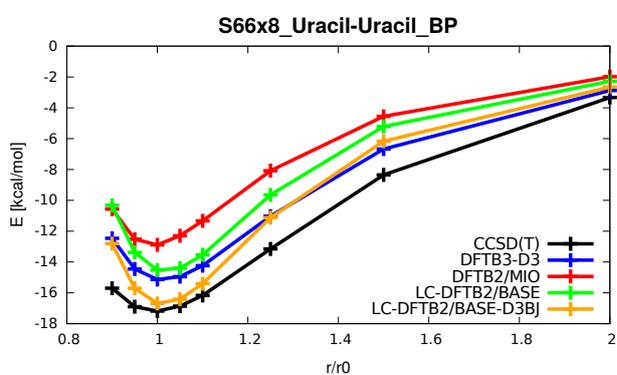
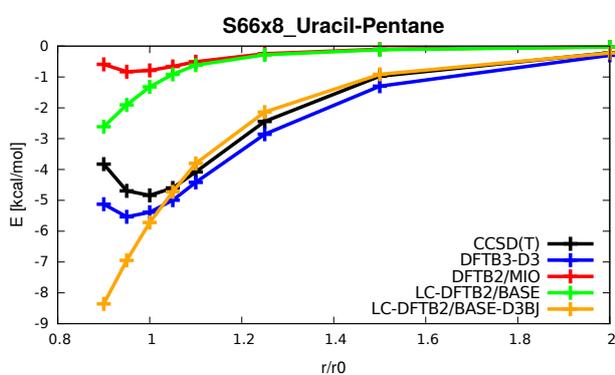
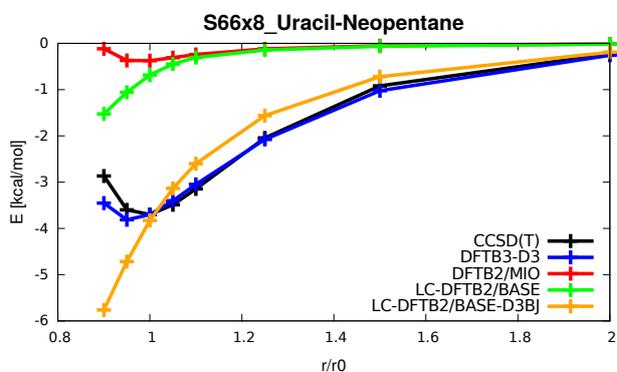
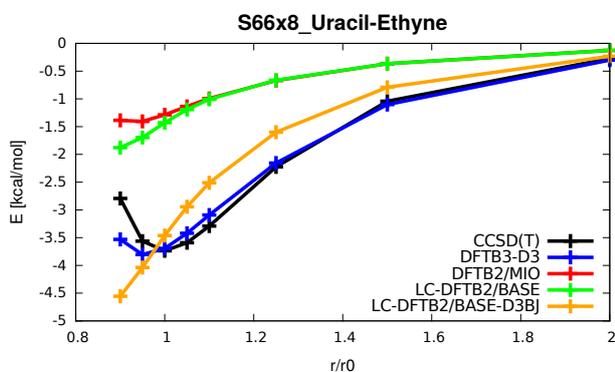


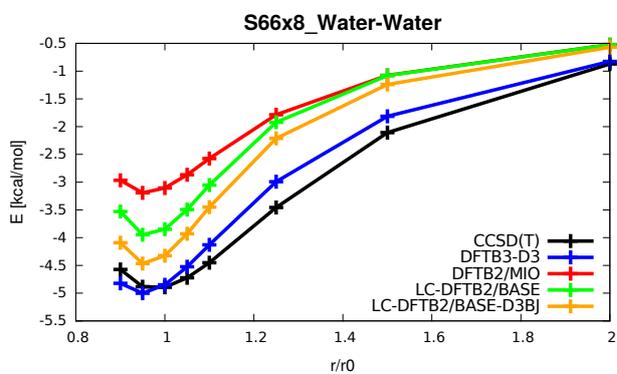
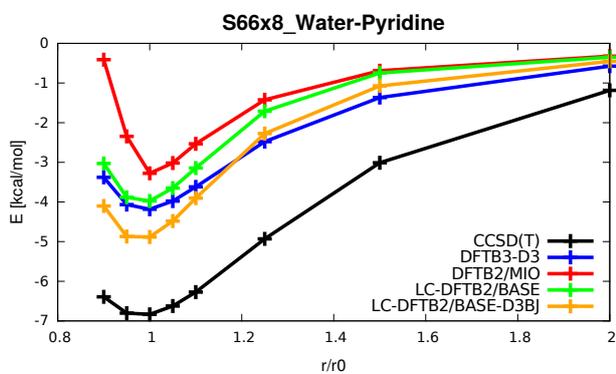












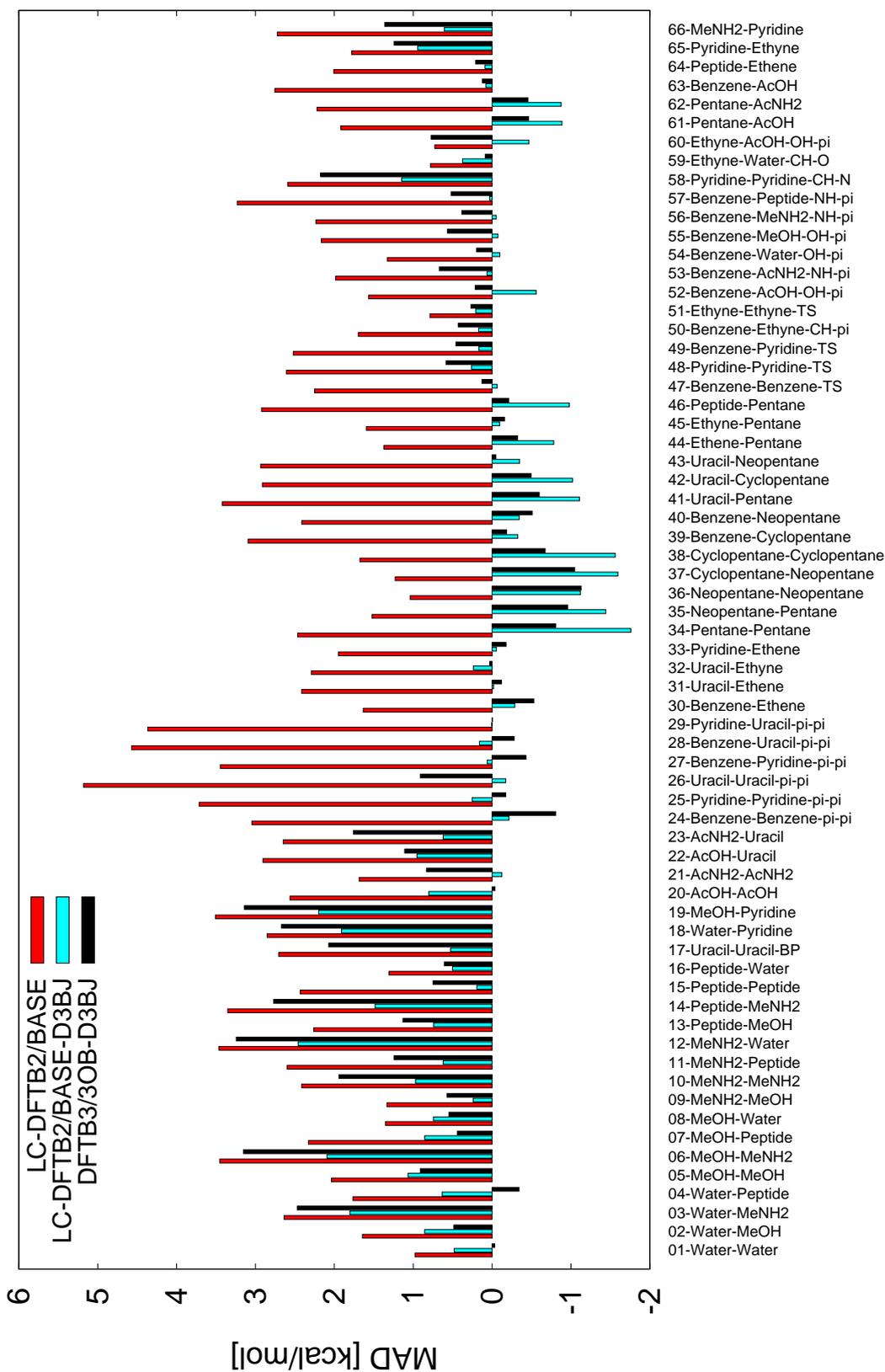


Figure C.1: Mean absolute deviations (in kcal/mol) of DFTB3/3OB-D3BJ (black), LC-DFTB2/BASE (red) and LC-DFTB2/BASE-D3BJ (cyan) for the S66 training set systems. The MAD of the non-dispersion corrected LC-DFTB2 amounts to 2.353 kcal/mol, while it is reduced with D3 correction to 0.667 kcal/mol.

Bibliography

- [1] Blankenship, R. E. Early Evolution of Photosynthesis. *Plant Physiol.* **2010**, *154*, 434–438.
- [2] Mitchell, P. Coupling of phosphorylation to electron and hydrogen transfer by a chemiosmotic type of mechanism. *Nature* **1961**, *191*, 144–148.
- [3] Woese, C. R.; Kandler, O.; Wheelis, M. L. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc. Nat. Acad. Sci. USA* **1990**, *87*, 4576–4579.
- [4] Oesterhelt, D.; Stoeckenius, W. Rhodopsin-like protein from the purple membrane of *Halobacterium halobium*. *Nat. New Biol.* **1971**, *233*, 149–152.
- [5] Stoeckenius, W.; Rowen, R. A morphological study of *Halobacterium halobium* and its lysis in media of low salt concentration. *J. Cell Biol.* **1967**, *34*, 365–393.
- [6] Oesterhelt, D.; Stoeckenius, W. Functions of a New Photoreceptor Membrane. *Proc. Nat. Acad. Sci. USA* **1973**, *70*, 2853–2857.
- [7] Oesterhelt, D.; Krippahl, G. Light inhibition of respiration in *Halobacterium halobium*. *FEBS Letters* **1973**, *36*, 72–76.
- [8] Danon, A.; Stoeckenius, W. Photophosphorylation in *Halobacterium halobium*. *Proc. Nat. Acad. Sci. USA* **1974**, *71*, 1234–8.
- [9] Hartmann, R.; Oesterhelt, D. Bacteriorhodopsin-Mediated Photophosphorylation in *Halobacterium halobium*. *Eur. J. Biochem.* **1977**, *77*, 325–335.
- [10] Racker, E.; Stoeckenius, W. Reconstitution of purple membrane vesicles catalyzing light driven proton uptake and adenosine triphosphate formation. *J. Biol. Chem.* **1974**, *249*, 662–663.
- [11] Henderson, R.; Unwin, P. N. Three-dimensional model of purple membrane obtained by electron microscopy. *Nature* **1975**, *257*, 28–32.
- [12] Neutze, R.; Pebay-Peyroula, E.; Edman, K.; Royant, A.; Navarro, J.; Landau, E. M. Bacteriorhodopsin: A high-resolution structural view of vectorial proton transport. *Biochim. Biophys. Acta, Biomembranes* **2002**, *1565*, 144–167.
- [13] Berman, H. M. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242.
- [14] Schobert, B.; Cupp-Vickery, J.; Hornak, V.; Smith, S. O.; Lanyi, J. K. Crystallographic structure of the K intermediate of bacteriorhodopsin: Conservation of free energy after photoisomerization of the retinal. *J. Mol. Biol.* **2002**, *321*, 715–726.
- [15] Lewis, A.; Spoonhower, J.; Bogomolni, R. a.; Lozier, R. H.; Stoeckenius, W. Tunable

- laser resonance Raman spectroscopy of bacteriorhodopsin. *Proc. Nat. Acad. Sci. USA* **1974**, *71*, 4462–4466.
- [16] Bayley, H.; Huang, K. S.; Radhakrishnan, R.; Ross, A. H.; Takagaki, Y.; Khorana, H. G. Site of attachment of retinal in bacteriorhodopsin. *Proc. Nat. Acad. Sci. USA* **1981**, *78*, 2225–9.
- [17] Lemke, H. D.; Oesterhelt, D. Lysine 216 is a binding site of the retinyl moiety in bacteriorhodopsin. *FEBS Letters* **1981**, *128*, 255–260.
- [18] Réat, V.; Patzelt, H.; Ferrand, M.; Pfister, C.; Oesterhelt, D.; Zaccai, G.; Reat, V.; Patzelt, H.; Ferrand, M.; Pfister, C.; Oesterhelt, D.; Zaccai, G. Dynamics of different functional parts of bacteriorhodopsin : H-2H labeling and neutron scattering. *Proc. Nat. Acad. Sci. USA* **1998**, *95*, 4970–4975.
- [19] Essen, L.; Siegert, R.; Lehmann, W. D.; Oesterhelt, D. Lipid patches in membrane protein oligomers: Crystal structure of the bacteriorhodopsin-lipid complex. *Proc. Nat. Acad. Sci. USA* **1998**, *95*, 11673–8.
- [20] Belrhali, H.; Nollert, P.; Royant, A.; Menzel, C.; Rosenbusch, J. P.; Landau, E. M.; Pebay-Peyroula, E. Protein, lipid and water organization in bacteriorhodopsin crystals: A molecular view of the purple membrane at 1.9 Å resolution. *Structure* **1999**, *7*, 909–917.
- [21] Luecke, H.; Schobert, B.; Richter, H. T.; Cartailler, J. P.; Lanyi, J. K. Structure of bacteriorhodopsin at 1.55 Å resolution. *J. Mol. Biol.* **1999**, *291*, 899–911.
- [22] Mathies, R. A. From femtoseconds to biology: Mechanism of bacteriorhodopsin's light-driven proton pump. *Proc. Indian Acad. Sci. - Chem. Sci.* **1991**, *103*, 283–293.
- [23] Lanyi, J. K. Bacteriorhodopsin. *Annu. Rev. Physiol.* **2004**, *66*, 665–688.
- [24] Bratanov, D.; Balandin, T.; Round, E.; Shevchenko, V.; Gushchin, I.; Polovinkin, V.; Borshchevskiy, V.; Gordeliy, V. An approach to heterologous expression of membrane proteins. The case of bacteriorhodopsin. *PLoS ONE* **2015**, *10*, e0128390.
- [25] Nango, E. et al. A three-dimensional movie of structural changes in bacteriorhodopsin. *Science* **2016**, *354*, 1552–1557.
- [26] Luecke, H.; Richter, H. T.; Lanyi, J. K. Proton transfer pathways in bacteriorhodopsin at 2.3 angstrom resolution. *Science* **1998**, *280*, 1934–1937.
- [27] Terentis, A. C.; Ujj, L.; Abramczyk, H.; Atkinson, G. H. Primary events in the bacteriorhodopsin photocycle: Torsional vibrational dephasing in the first excited electronic state. *Chem. Phys.* **2005**, *313*, 51–62.
- [28] Brack, T. L.; Atkinson, G. H. Picosecond time-resolved resonance Raman spectrum of the K-590 intermediate in the room temperature bacteriorhodopsin photocycle. *J. Mol. Struct.* **1989**, *214*, 289–303.
- [29] Sharkov, A. V.; Pakulev, A. V.; Chekalin, S. V.; Matveetz, Y. A. Primary events in bacteriorhodopsin probed by subpicosecond spectroscopy. *Biochim. Biophys. Acta, Bioenerg.* **1985**, *808*, 94–102.
- [30] Polland, H. J.; Franz, M. A.; Zinth, W.; Kaiser, W.; Kölling, E.; Oesterhelt, D. Early Picosecond Events in the Photocycle of Bacteriorhodopsin. *Biophys. J* **1986**,

- 49, 651–662.
- [31] Edman, K.; Nollert, P.; Royant, A.; Belrhali, H.; Pebay-Peyroula, E.; Hajdu, J.; Neutze, R.; Landau, E. M. High-resolution X-ray structure of an early intermediate in the bacteriorhodopsin photocycle. *Nature* **1999**, *401*, 822–826.
- [32] Stoeckenius, W.; Lozier, R. H. Light energy conversion in *Halobacterium halobium*. *J. Supramol. Struct.* **1974**, *2*, 769–774.
- [33] Nagle, J. F.; Parodi, L. A.; Lozier, R. H. Procedure for testing kinetic models of the photocycle of bacteriorhodopsin. *Biophys. J* **1982**, *38*, 161–174.
- [34] Lozier, R. H.; Bogomolni, R. A.; Stoeckenius, W. Bacteriorhodopsin: a light-driven proton pump in *Halobacterium Halobium*. *Biophys. J.* **1975**, *15*, 955–962.
- [35] Bondar, A. N.; Elstner, M.; Suhai, S.; Smith, J. C.; Fischer, S. Mechanism of primary proton transfer in bacteriorhodopsin. *Structure* **2004**, *12*, 1281–1288.
- [36] Gerwert, K.; Souvignier, G.; Hess, B. Simultaneous monitoring of light-induced changes in protein side-group protonation, chromophore isomerization, and backbone motion of bacteriorhodopsin by time-resolved Fourier-transform infrared spectroscopy. *Proc. Nat. Acad. Sci. USA* **1990**, *87*, 9774–8.
- [37] Ames, J. B.; Fodor, S. P.; Gebhard, R.; Raap, J.; van den Berg, E. M.; Lugtenburg, J.; Mathies, R. A. Bacteriorhodopsin's M412 intermediate contains a 13-cis, 14-s-trans, 15-anti-retinal Schiff base chromophore. *Biochemistry* **1989**, *28*, 3681–3687.
- [38] Aton, B.; Doukas, A. G.; Callender, R. H.; Becker, B.; Ebrey, T. G. Resonance Raman Studies of the Purple Membrane. *Biochemistry* **1977**, *16*, 2995–2999.
- [39] Ludmann, K.; Gergely, C.; Váró, G. Kinetic and thermodynamic study of the bacteriorhodopsin photocycle over a wide pH range. *Biophys. J* **1998**, *75*, 3110–3119.
- [40] Luecke, H. Atomic resolution structures of bacteriorhodopsin photocycle intermediates: The role of discrete water molecules in the function of this light-driven ion pump. *Biochim. Biophys. Acta, Bioenergetics* **2000**, *1460*, 133–156.
- [41] Luecke, H.; Schobert, B.; Richter, H. T.; Cartailler, J. P.; Lanyi, J. K. Structural changes in bacteriorhodopsin during ion transport at 2 angstrom resolution. *Science* **1999**, *286*, 255–260.
- [42] Luecke, H.; Schobert, B.; Cartailler, J. P.; Richter, H. T.; Rosengarth, A.; Needleman, R.; Lanyi, J. K. Coupling photoisomerization of retinal to directional transport in bacteriorhodopsin. *J. Mol. Biol.* **2000**, *300*, 1237–1255.
- [43] Lanyi, J. K.; Schobert, B. Structural Changes in the L Photointermediate of Bacteriorhodopsin. *J. Mol. Biol.* **2007**, *365*, 1379–1392.
- [44] Clemens, M.; Phatak, P.; Cui, Q.; Bondar, A. N.; Elstner, M. Role of Arg82 in the early steps of the bacteriorhodopsin proton-pumping cycle. *J. Phys. Chem. B* **2011**, *115*, 7129–7135.
- [45] Brown, L. S.; Sasaki, J.; Kandori, H.; Maeda, A.; Needleman, R.; Lanyi, J. K. Glutamic acid 204 is the terminal proton release group at the extracellular surface of bacteriorhodopsin. *J. Biol. Chem.* **1995**, *270*, 27122–27126.
- [46] Balashov, S. P.; Imasheva, E. S.; Ebrey, T. G.; Chen, N.; Menick, D. R.; Crouch, R. K.

- Glutamate-194 to cysteine mutation inhibits fast light-induced proton release in bacteriorhodopsin. *Biochemistry* **1997**, *36*, 8671–8676.
- [47] Garczarek, F.; Gerwert, K. Functional waters in intraprotein proton transfer monitored by FTIR difference spectroscopy. *Nature* **2006**, *439*, 109–112.
- [48] Phatak, P.; Ghosh, N.; Yu, H.; Cui, Q.; Elstner, M. Amino acids with an intermolecular proton bond as proton storage site in bacteriorhodopsin. *Proc. Nat. Acad. Sci. USA* **2008**, *105*, 19672–19677.
- [49] Goyal, P.; Ghosh, N.; Phatak, P.; Clemens, M.; Gaus, M.; Elstner, M.; Cui, Q. Proton storage site in bacteriorhodopsin: New insights from quantum mechanics/molecular mechanics simulations of microscopic p K a and infrared spectra. *J. Am. Chem. Soc.* **2011**, *133*, 14981–14997.
- [50] Hessling, B.; Herbst, J.; Rammelsberg, R.; Gerwert, K. Fourier transform infrared double-flash experiments resolve bacteriorhodopsin's M1 to M2 transition. *Biophys. J* **1997**, *73*, 2071–2080.
- [51] Subramaniam, S.; Lindahl, M.; Bullough, P.; Faruqi, A. R.; Tittor, J.; Oesterhelt, D.; Brown, L.; Lanyi, J.; Henderson, R. Protein conformational changes in the bacteriorhodopsin photocycle. *J. Mol. Biol.* **1999**, *287*, 145–161.
- [52] Bousché, O.; Braiman, M.; He, Y. W.; Marti, T.; Khorana, H. G.; Rothschild, K. J. Vibrational spectroscopy of bacteriorhodopsin mutants: Evidence that Asp-96 deprotonates during the M N transition. *J. Biol. Chem.* **1991**, *266*, 11063–11067.
- [53] Holz, M.; Drachev, L. A.; Mogi, T.; Otto, H.; Kaulen, A. D.; Heyn, M. P.; Skulachev, V. P.; Khorana, H. G. Replacement of aspartic acid-96 by asparagine in bacteriorhodopsin slows both the decay of the M intermediate and the associated proton movement. *Proc. Nat. Acad. Sci. USA* **1989**, *86*, 2167–2171.
- [54] Souvignier, G.; Gerwert, K. Proton uptake mechanism of bacteriorhodopsin as determined by time-resolved stroboscopic-FTIR-spectroscopy. *Biophys. J* **1992**, *63*, 1393–1405.
- [55] Otto, H.; Marti, T.; Holz, M.; Mogi, T.; Lindau, M.; Khorana, H. G.; Heyn, M. P. Aspartic acid-96 is the internal proton donor in the reprotonation of the Schiff base of bacteriorhodopsin. *Proc. Nat. Acad. Sci. USA* **1989**, *86*, 9228–32.
- [56] Gerwert, K.; Hess, B.; Soppa, J.; Oesterhelt, D. Role of aspartate-96 in proton translocation by bacteriorhodopsin. *Proc. Nat. Acad. Sci. USA* **1989**, *86*, 4943–7.
- [57] Cao, Y.; Váró, G.; Klinger, A. L.; Czajkowsky, D. M.; Braiman, M. S.; Needleman, R.; Lanyi, J. K. Proton transfer from asp-96 to the bacteriorhodopsin Schiff base is caused by a decrease of the pKa of asp-96 which follows a protein backbone conformational change. *Biochemistry* **1993**, *32*, 1981–1990.
- [58] Miller, A.; Oesterhelt, D. Kinetic optimization of bacteriorhodopsin by aspartic acid 96 as an internal proton donor. *Biochim. Biophys. Acta, Bioenerg.* **1990**, *1020*, 57–64.
- [59] Schobert, B.; Brown, L. S.; Lanyi, J. K. Crystallographic structures of the M and N intermediates of bacteriorhodopsin: Assembly of a hydrogen-bonded chain of water molecules between Asp-96 and the retinal Schiff base. *J. Mol. Biol.* **2003**, *330*, 553–570.

- [60] Gai, F.; Hasson, K. C.; McDonald, J. C.; Anfinrud, P. A. Chemical dynamics in proteins: The photoisomerization of retinal in bacteriorhodopsin. *Science* **1998**, *279*, 1886–1891.
- [61] Dioumaev, A. K.; Brown, L. S.; Needleman, R.; Lanyi, J. K. Partitioning of free energy gain between the photoisomerized retinal and the protein in bacteriorhodopsin. *Biochemistry* **1998**, *37*, 9889–9893.
- [62] Smith, S. O.; Curry, B.; Mathies, R.; Pardoën, J. A.; Mulder, P. P.; Lugtenburg, J. Chromophore Structure in Bacteriorhodopsin's O640Photointermediate. *Biochemistry* **1983**, *22*, 6141–6148.
- [63] Ames, J.; Biochemistry, R. M.; 1990, U. The role of back-reactions and proton uptake during the N. fwardw. O transition in bacteriorhodopsin's photocycle: a kinetic resonance Raman study. *Biochemistry* **1990**,
- [64] Richter, H. T.; Needleman, R.; Kandori, H.; Maeda, A.; Lanyi, J. K. Relationship of retinal configuration and internal proton transfer at the end of the bacteriorhodopsin photocycle. *Biochemistry* **1996**, *35*, 15461–15466.
- [65] Dioumaev, A. K.; Brown, L. S.; Needleman, R.; Lanyi, J. K. Fourier transform infrared spectra of a late intermediate of the bacteriorhodopsin photocycle suggest transient protonation of Asp-212. *Biochemistry* **1999**, *38*, 10070–10078.
- [66] Zscherp, C.; Schlesinger, R.; Heberle, J. Time-resolved FT-IR spectroscopic investigation of the pH-dependent proton transfer reactions in the E194Q mutant of bacteriorhodopsin. *Biochem. Biophys. Res. Commun.* **2001**, *283*, 57–63.
- [67] Phatak, P.; Frähmcke, J. S.; Wanko, M.; Hoffmann, M.; Strodel, P.; Smith, J. C.; Suhai, S.; Bondar, A. N.; Elstner, M. Long-distance proton transfer with a break in the bacteriorhodopsin active site. *J. Am. Chem. Soc.* **2009**, *131*, 7064–7078.
- [68] Phatak, P. V. Investigation of Proton Transfer Pathways in Bacteriorhodopsin with Multi-Length-Scale Simulations. Ph.D. thesis, TU Braunschweig, 2009.
- [69] Ge, X.; Gunner, M. R. Unraveling the mechanism of proton translocation in the extracellular half-channel of bacteriorhodopsin. *Proteins Struct. Funct. Bioinforma.* **2016**, *84*, 639–654.
- [70] Imasheva, E. S.; Balashov, S. P.; Ebrey, T. G.; Chen, N.; Crouch, R. K.; Menick, D. R. Two groups control light-induced Schiff base deprotonation and the proton affinity of Asp85in the Arg82his mutant of bacteriorhodopsin. *Biophys. J* **1999**, *77*, 2750–2763.
- [71] Orto, H.; Martit, T.; Holz, M.; Mogitt, T.; Sternt, L. J.; Engel, F.; Khoranat, H. G.; Heyn, M. P. Substitution of amino acids Asp-85, Asp-212, and Arg-82 in bacteriorhodopsin affects the proton release phase of the pump and the pK of the Schiff base. *Biophysics* **1990**, *87*, 1018–1022.
- [72] Rouhani, S.; Cartailier, J. P.; Facciotti, M. T.; Walian, P.; Needleman, R.; Lanyi, J. K.; Glaeser, R. M.; Luecke, H. Crystal structure of the D85S mutant of bacteriorhodopsin: Model of an O-like photocycle intermediate. *J. Mol. Biol.* **2001**, *313*, 615–628.
- [73] Oesterhelt, D. The structure and mechanism of the family of retinal proteins from halophilic archaea. *Curr. Opin. Struct. Biol.* **1998**, *8*, 489–500.

- [74] Elstner, M.; Seifert, G. Density functional tight binding. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **2014**, *372*, 20120483–20120483.
- [75] Seifert, G.; Porezag, D.; Frauenheim, T. Calculations of molecules, clusters, and solids with a simplified LCAO-DFT-LDA scheme. *Int. J. Quantum Chem.* **1996**, *58*, 185–192.
- [76] Porezag, D.; Frauenheim, T.; Köhler, T.; Seifert, G.; Kaschner, R. Construction of tight-binding-like potentials on the basis of density-functional theory: Application to carbon. *Phys. Rev. B* **1995**, *51*, 12947–12957.
- [77] Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties. *Phys. Rev. B* **1998**, *58*, 7260–7268.
- [78] Yang, Y.; Yu, H.; York, D.; Cui, Q.; Elstner, M. Extension of the self-consistent-charge density-functional tight-binding method: Third-order expansion of the density functional theory total energy and introduction of a modified effective coulomb interaction. *J. Phys. Chem. A* **2007**, *111*, 10861–10873.
- [79] Elstner, M. SCC-DFTB: What is the proper degree of self-consistency. *J. Phys. Chem. A* **2007**, *111*, 5614–5621.
- [80] Gaus, M.; Cui, Q.; Elstner, M. DFTB3: Extension of the self-consistent-charge density-functional tight-binding method (SCC-DFTB). *J. Chem. Theory Comput.* **2011**, *7*, 931–948.
- [81] Gaus, M.; Goez, A.; Elstner, M. Parametrization and benchmark of DFTB3 for organic molecules. *J. Chem. Theory Comput.* **2013**, *9*, 338–354.
- [82] Elstner, M. The SCC-DFTB method and its application to biological systems. *Theor. Chem. Acc.* **2006**, *116*, 316–325.
- [83] Goyal, P.; Yang, S.; Cui, Q. Microscopic basis for kinetic gating in cytochrome c oxidase: insights from QM/MM analysis. *Chem. Sci.* **2015**, *6*, 826–841.
- [84] Cui, Q.; Karplus, M. Is a "proton wire" concerted or stepwise? A model study of proton transfer in carbonic anhydrase. *J. Phys. Chem. B* **2003**, *107*, 1071–1078.
- [85] Smedarchina, Z.; Siebrand, W.; Fernández-Ramos, A.; Cui, Q. Kinetic isotope effects for concerted multiple proton transfer: A direct dynamics study of an active-site model of carbonic anhydrase II. *J. Am. Chem. Soc.* **2003**, *125*, 243–251.
- [86] Riccardi, D.; König, P.; Guo, H.; Cui, Q. Proton transfer in carbonic anhydrase is controlled by electrostatics rather than the orientation of the acceptor. *Biochemistry* **2008**, *47*, 2369–2378.
- [87] Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.
- [88] Grimme, S.; Hansen, A.; Brandenburg, J. G.; Bannwarth, C. Dispersion-Corrected Mean-Field Electronic Structure Methods. *Chem. Rev.* **2016**, *116*, 5105–5154.
- [89] Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **2010**, *132*.

- [90] Brandenburg, J. G.; Grimme, S. Accurate modeling of organic molecular crystals by dispersion-corrected density functional tight binding (DFTB). *J. Phys. Chem. Letters* **2014**, *5*, 1785–1789.
- [91] Kubillus, M.; Kubař, T.; Gaus, M.; Řezáč, J.; Elstner, M. Parameterization of the DFTB3 method for Br, Ca, Cl, F, I, K, and Na in organic and biological systems. *J. Chem. Theory Comput.* **2015**, *11*, 332–342.
- [92] Christensen, A. S.; Kubař, T.; Cui, Q.; Elstner, M. Semiempirical Quantum Mechanical Methods for Noncovalent Interactions for Chemical and Biochemical Applications. *Chem. Rev.* **2016**, *116*, 5301–5337.
- [93] Liang, R.; Swanson, J. M. J.; Voth, G. A. Benchmark study of the SCC-DFTB approach for a biomolecular proton channel. *J. Chem. Theory Comput.* **2014**, *10*, 451–462.
- [94] Goyal, P.; Elstner, M.; Cui, Q. Application of the SCC-DFTB method to neutral and protonated water clusters and bulk water. *J. Phys. Chem. B* **2011**, *115*, 6790–6805.
- [95] Goyal, P.; Qian, H. J.; Irle, S.; Lu, X.; Roston, D.; Mori, T.; Elstner, M.; Cui, Q. Molecular simulation of water and hydration effects in different environments: Challenges and developments for DFTB based models. *J. Phys. Chem. B* **2014**, *118*, 11007–11027.
- [96] Choi, T. H.; Liang, R.; Maupin, C. M.; Voth, G. A. Application of the SCC-DFTB method to hydroxide water clusters and aqueous hydroxide solutions. *J. Phys. Chem. B* **2013**, *117*, 5165–5179.
- [97] Kubillus, M. Delta-Pauli Theory and Implementation in DFTB+ (Unpublished). Ph.D. thesis, Karlsruhe Institute for Technology (KIT).
- [98] Řezáč, J.; Riley, K. E.; Hobza, P. S66: A well-balanced database of benchmark interaction energies relevant to biomolecular structures. *J. Chem. Theory Comput.* **2011**, *7*, 2427–2438.
- [99] Born, M.; Oppenheimer, R. Zur Quantentheorie der Molekeln. *Ann. Phys.* **1927**, *389*, 457–484.
- [100] Hohenberg, P.; Kohn, W. Inhomogeneous Electron Gas. *Phys. Rev.* **1964**, *136*, B864–B871.
- [101] Thomas, L. H. The calculation of atomic fields. *Math. Proc. Cambridge Philos. Soc.* **1927**, *23*, 542–548.
- [102] Fermi, E. Statistical method to determine some properties of atoms. *Rend. Accad. Naz. Lincei* **1927**, *6*, 602–607.
- [103] Dirac, P. A. M. Note on Exchange Phenomena in the Thomas Atom. *Math. Proc. Cambridge Philos. Soc.* **1930**, *26*, 376–385.
- [104] Kohn, W.; Sham, L. J. Self-consistent equations including exchange and correlation effects. *Phys. Rev.* **1965**, *140*, A1133–A1138.
- [105] Perdew, J. P.; Wang, Y. Accurate and simple analytic representation of the electron-gas correlation energy. *Phys. Rev. B* **1992**, *45*, 13244–13249.
- [106] Becke, A. D. Density-functional exchange-energy approximation with correct asymp-

- otic behavior. *Phys. Rev. A* **1988**, *38*, 3098–3100.
- [107] Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **1988**, *37*, 785–789.
- [108] Tao, J.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. Climbing the Density Functional Ladder: Non-Empirical Meta-Generalized Gradient Approximation Designed for Molecules and Solids. *Phys. Rev. Lett.* **2003**, *91*, 146401.
- [109] Becke, A. D. Densityfunctional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648–5652.
- [110] Vosko, S. H.; Wilk, L.; Nusair, M. Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. *Can. J. Phys.* **1980**, *58*, 1200–1211.
- [111] Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. Ab Initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields. *J. Phys. Chem.* **1994**, *98*, 11623–11627.
- [112] Foulkes, W. M. C.; Haydock, R. Tight-binding models and density-functional theory. *Phys. Rev. B* **1989**, *39*, 12520–12536.
- [113] Eschrig, H. *Optimized LCAO Method and the Electronic Structure of Extended Systems*; Research Reports in Physics; Springer Berlin Heidelberg: Berlin, Heidelberg, 1989; p 221.
- [114] Ohno, K. Some remarks on the Pariser-Parr-Pople method. *Theor. Chim. Acta* **1964**, *2*, 219–227.
- [115] Klopman, G. A Semiempirical Treatment of Molecular Structures. III. Equipotential Orbitals for Polyatomic Systems. *J. Am. Chem. Soc.* **1965**, *87*, 3300–3303.
- [116] Johnson, E. R.; Becke, A. D. A post-Hartree-Fock model of intermolecular interactions: Inclusion of higher-order corrections. *J. Chem. Phys.* **2006**, *124*, 174104.
- [117] Johnson, E. R.; Becke, A. D. A post-Hartree-Fock model of intermolecular interactions. *J. Chem. Phys.* **2005**, *123*, 024101.
- [118] Becke, A. D.; Johnson, E. R. A density-functional model of the dispersion interaction. *J. Chem. Phys.* **2005**, *123*, 154101.
- [119] Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* **2011**, *32*, 1456–1465.
- [120] Giese, T. J.; Chen, H.; Dissanayake, T.; Giambau, G. M.; Heldenbrand, H.; Huang, M.; Kuechler, E. R.; Lee, T. S.; Panteva, M. T.; Radak, B. K.; York, D. M. A variational linear-scaling framework to build practical, efficient next-generation orbital-based quantum force fields. *J. Chem. Theory Comput.* **2013**, *9*, 1417–1427.
- [121] Van Vleet, M. J.; Misquitta, A. J.; Stone, A. J.; Schmidt, J. R. Beyond Born-Mayer: Improved Models for Short-Range Repulsion in ab Initio Force Fields. *J. Chem. Theory Comput.* **2016**, *12*, 3851–3870.
- [122] Marini, F.; Walczak, B. Particle swarm optimization (PSO). A tutorial. *Chemom. Intell. Lab. Syst.* **2015**, *149*, 153–165.
- [123] Vuong, Q. V.; Akkarapattiakal Kuriappan, J.; Kubillus, M.; Kranz, J.; Mast, T.;

- Niehaus, T. A.; Irle, S.; Elstner, M. Parametrization and Benchmark of the Long-range Corrected DFTB2 for Organic Molecules. *J. Chem. Theory Comput.* **2017**, *14*, 115–125.
- [124] Christensen, A. S.; Elstner, M.; Cui, Q. Improving intermolecular interactions in DFTB3 using extended polarization from chemical-potential equalization. *J. Chem. Phys.* **2015**, *143*, 084123.
- [125] Chou, C. P.; Nishimura, Y.; Fan, C. C.; Mazur, G.; Irle, S.; Witek, H. a. Automated Parameterization of DFTB Using Particle Swarm Optimization. *J. Chem. Theory Comput.* **2016**, *12*, 53–64.
- [126] Eberhart, R.; Kennedy, J. A new optimizer using particle swarm theory. Proceedings of the Sixth International Symposium on Micro Machine and Human Science MHS'95. 1995; pp 39–43.
- [127] Shi, Y.; Eberhart, R. A modified particle swarm optimizer. IEEE International Conference on Evolutionary Computation Proceedings. 1998; pp 69–73.
- [128] Zeugmann, T. et al. In *Encyclopedia of Machine Learning*; Sammut, C., Webb, G. I., Eds.; Springer US: Boston, MA, 2011; pp 760–766.
- [129] Eberhart, R.; Yuhui Shi, Particle swarm optimization: developments, applications and resources. Proceedings of the Congress on Evolutionary Computation (IEEE Cat. No.01TH8546). 2001; pp 81–86.
- [130] Lindahl, E.; Bjelkmar, P.; Larsson, P.; Cuendet, M. A.; Hess, B. Implementation of the charmm force field in GROMACS: Analysis of protein stability effects from correction maps, virtual interaction sites, and water models. *J. Chem. Theory Comput.* **2010**, *6*, 459–466.
- [131] Klauda, J. B.; Venable, R. M.; Freites, J. A.; O'Connor, J. W.; Tobias, D. J.; Mondragon-Ramirez, C.; Vorobyov, I.; MacKerell, A. D.; Pastor, R. W. Update of the CHARMM All-Atom Additive Force Field for Lipids: Validation on Six Lipid Types. *J. Phys. Chem. B* **2010**, *114*, 7830–7843.
- [132] Mackerell, A. D.; Feig, M.; Brooks, C. L. Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulation. *J. Comput. Chem.* **2004**, *25*, 1400–1415.
- [133] Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577–8593.
- [134] Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
- [135] Hockney, R. W.; Goel, S. P.; Eastwood, J. W. Quiet high-resolution computer models of a plasma. *J. Comput. Phys.* **1974**, *14*, 148–158.
- [136] Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977**, *23*, 327–341.
- [137] Hess, B.; Bekker, H.; Berendsen, H. J.; Fraaije, J. G. LINCS: A Linear Constraint Solver for molecular simulations. *J. Comput. Chem.* **1997**, *18*, 1463–1472.

- [138] Hoover, W. G. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A* **1985**, *31*, 1695–1697.
- [139] Nosé, S. A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.* **1984**, *81*, 511–519.
- [140] Nosé, S. A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.* **1984**, *52*, 255–268.
- [141] Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981**, *52*, 7182–7190.
- [142] König, P. H.; Hoffmann, M.; Frauenheim, T.; Cui, Q. A critical evaluation of different QM/MM frontier treatments with SCC-DFTB as the QM method. *J. Phys. Chem. B* **2005**, *109*, 9082–9095.
- [143] Neumann, J. V. Physical Applications of the Ergodic Hypothesis. *Proc. Nat. Acad. Sci. USA* **1932**, *18*, 263–6.
- [144] von Neumann, J. Proof of the Quasi-ergodic Hypothesis. *Proc. Nat. Acad. Sci. USA* **1932**, *18*, 70–82.
- [145] Birkhoff, G. D. Proof of the Ergodic Theorem. *Proc. Nat. Acad. Sci. USA* **1931**, *17*, 656–660.
- [146] Ode, H.; Nakashima, M.; Kitamura, S.; Sugiura, W.; Sato, H. Molecular dynamics simulation in virus research. *Front. Microbiol.* **2012**, *3*, 258.
- [147] Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. Equation of state calculations by fast computing machines. *J. Chem. Phys.* **1953**, *21*, 1087–1092.
- [148] Sugita, Y.; Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Letters* **1999**, *314*, 141–151.
- [149] Roe, D. R.; Bergonzo, C.; Cheatham, T. E. Evaluation of enhanced sampling provided by accelerated molecular dynamics with hamiltonian replica exchange methods. *J. Phys. Chem. B* **2014**, *118*, 3543–3552.
- [150] Ostermeir, K.; Zacharias, M. Advanced replica-exchange sampling to study the flexibility and plasticity of peptides and proteins. *Biochim. Biophys. Acta, Proteins and Proteomics* **2013**, *1834*, 847–853.
- [151] Bussi, G. Hamiltonian replica exchange in GROMACS: A flexible implementation. *Mol. Phys.* **2014**, *112*, 379–384.
- [152] Kirkwood, J. G. Statistical mechanics of fluid mixtures. *J. Chem. Phys.* **1935**, *3*, 300–313.
- [153] Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. THE weighted histogram analysis method for freeenergy calculations on biomolecules. I. The method. *J. Comput. Chem.* **1992**, *13*, 1011–1021.
- [154] Laio, A.; Parrinello, M. Escaping free-energy minima. *Proc. Nat. Acad. Sci. USA* **2002**, *99*, 12562–12566.
- [155] Barducci, A.; Bussi, G.; Parrinello, M. Well-tempered metadynamics: A smoothly converging and tunable free-energy method. *Phys. Rev. Lett.* **2008**, *100*, 020603.

- [156] Laio, A.; Rodriguez-Fortea, A.; Gervasio, F. L.; Ceccarelli, M.; Parrinello, M. Assessing the accuracy of metadynamics. *J. Phys. Chem. B* **2005**, *109*, 6714–6721.
- [157] Raiteri, P.; Laio, A.; Gervasio, F. L.; Micheletti, C.; Parrinello, M. Efficient reconstruction of complex free energy landscapes by multiple walkers metadynamics. *J. Phys. Chem. B* **2006**, *110*, 3533–3539.
- [158] Smondyrev, A. M.; Voth, G. A. Molecular Dynamics Simulation of Proton Transport through the Influenza A Virus M2 Channel. *Biophys. J* **2002**, *83*, 1987–1996.
- [159] König, P. H.; Ghosh, N.; Hoffmann, M.; Elstner, M.; Tajkhorshid, E.; Frauenheim, T.; Cui, Q. Toward theoretical analysis of long-range proton transfer kinetics in biomolecular pumps. *J. Phys. Chem. A* **2006**, *110*, 548–563.
- [160] Watanabe, H. C.; Ishikura, T.; Yamato, T. Theoretical modeling of the O-intermediate structure of bacteriorhodopsin. *Proteins Struct. Funct. Bioinforma.* **2009**, *75*, 53–61.
- [161] Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindah, E. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1-2*, 19–25.
- [162] Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New feathers for an old bird. *Comput. Phys. Commun.* **2014**, *185*, 604–613.
- [163] Baudry, J.; Crouzy, S.; Roux, B.; Smith, J. C. Quantum chemical and free energy simulation analysis of retinal conformational energetics. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 1018–1024.
- [164] Nina, M.; Roux, B.; Smith, J. C. Functional interactions in bacteriorhodopsin: a theoretical analysis of retinal hydrogen bonding with water. *Biophys. J* **1995**, *68*, 25–39.
- [165] Tajkhorshid, E.; Baudry, J.; Schulten, K.; Suhai, S. Molecular dynamics study of the nature and origin of retinal's twisted structure in bacteriorhodopsin. *Biophys. J.* **2000**, *78*, 683–693.
- [166] Tajkhorshid, E.; Suhai, S. Influence of the Methyl Groups on the Structure, Charge Distribution, and Proton Affinity of the Retinal Schiff Base. *J. Phys. Chem. B* **1999**, *103*, 5581–5590.
- [167] Tajkhorshid, E.; Paizs, B.; Suhai, S. Conformational Effects on the Proton Affinity of the Schiff Base in Bacteriorhodopsin: A Density Functional Study. *J. Phys. Chem. B* **1997**, *101*, 8021–8028.
- [168] Kandt, C.; Ash, W. L.; Peter Tieleman, D. Setting up and running molecular dynamics simulations of membrane proteins. *Methods* **2007**, *41*, 475–488.
- [169] Tieleman, D. Lipid properties and the orientation of aromatic residues in OmpF, influenza M2, and alamethicin systems: Molecular dynamics simulations. *Biochemistry* **1998**, *37*, 17554–17561.
- [170] Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.
- [171] Grossfield, A. WHAM: the weighted histogram analysis method. <http://membrane>.

- urmc.rochester.edu/content/wham.
- [172] Armstrong, C. T.; Mason, P. E.; Anderson, J. L. R.; Dempsey, C. E. Arginine side chain interactions and the role of arginine as a gating charge carrier in voltage sensitive ion channels. *Sci. Rep.* **2016**, *6*, 21759.
- [173] Kubař, T.; Welke, K.; Groenhof, G. New QM/MM implementation of the DFTB3 method in the gromacs package. *J. Comput. Chem.* **2015**, *36*, 1978–1989.
- [174] Goerigk, L.; Grimme, S. A thorough benchmark of density functional methods for general main group thermochemistry, kinetics, and noncovalent interactions. *Phys. Chem. Chem. Phys.* **2011**, *13*, 6670.
- [175] Berg, J. M.; Tymoczko, J. L.; Stryer, L. *Stryer Biochemie*; Springer Berlin Heidelberg: Berlin, Heidelberg, 2013.
- [176] Li, L.; Li, C.; Zhang, Z.; Alexov, E. On the dielectric "constant" of proteins: Smooth dielectric function for macromolecular modeling and its implementation in DelPhi. *J. Chem. Theory Comput.* **2013**, *9*, 2126–2136.
- [177] Neese, F. The ORCA program system. *WIREs Comput. Mol. Sci.* **2012**, *2*, 73–78.
- [178] Kaila, V. R. I.; Hummer, G. Energetics and dynamics of proton transfer reactions along short water wires. *Phys. Chem. Chem. Phys.* **2011**, *13*, 13207.
- [179] Weigend, F. Accurate Coulomb-fitting basis sets for H to Rn. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1057.
- [180] Weigend, F.; Ahlrichs, R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297.
- [181] Izsák, R.; Neese, F. An overlap fitted chain of spheres exchange method. *J. Chem. Phys.* **2011**, *135*, 144105.
- [182] Neese, F.; Wennmohs, F.; Hansen, A.; Becker, U. Efficient, approximate and parallel Hartree-Fock and hybrid DFT calculations. A 'chain-of-spheres' algorithm for the Hartree-Fock exchange. *Chem. Phys.* **2009**, *356*, 98–109.
- [183] Baerends, E. J.; Ellis, D. E.; Ros, P. Self-consistent molecular Hartree-Fock-Slater calculations I. The computational procedure. *Chem. Phys.* **1973**, *2*, 41–51.
- [184] Dunlap, B. I.; Connolly, J. W. D.; Sabin, J. R. On some approximations in applications of $X\alpha$ theory. *J. Chem. Phys.* **1979**, *71*, 3396.
- [185] Vahtras, O.; Almlöf, J.; Feyereisen, M. W. Integral approximations for LCAO-SCF calculations. *Chem. Phys. Letters* **1993**, *213*, 514–518.
- [186] Herzfeld, K. F. Zur Theorie der Reaktionsgeschwindigkeiten in Gasen. *Ann. Phys.* **1919**, *364*, 635–667.
- [187] MacKerell, A. D. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **1998**, *102*, 3586.
- [188] Aradi, B.; Hourahine, B.; Frauenheim, T. DFTB+, a sparse matrix-based implementation of the DFTB method. *J. Phys. Chem. A* **2007**, *111*, 5678–5684.
- [189] Řezáč, J.; Jurečka, P.; Riley, K. E.; Černý, J.; Valdes, H.; Pluháčková, K.; Berka, K.; Řezáč, T.; Pitoák, M.; Vondrášek, J.; Hobza, P. Quantum Chemical Benchmark En-

- ergy and Geometry Database for Molecular Clusters and Complex Molecular Systems (www.begdb.com): A Users Manual and Examples. *Collect. Czechoslov. Chem. Commun.* **2008**, *73*, 1261–1270.
- [190] Frisch, M. J. et al. Gaussian 09, Revision A.02. 2009.
- [191] TURBOMOLE-GmbH, TURBOMOLE V7.0 2015, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH. 2015.
- [192] Takahashi, O.; Kohno, Y.; Nishio, M. Relevance of weak hydrogen bonds in the conformation of organic compounds and bioconjugates: Evidence from recent experimental data and high-level ab initio MO calculations. *Chem. Rev.* **2010**, *110*, 6049–6076.
- [193] Manna, D.; Kesharwani, M. K.; Sylvetsky, N.; Martin, J. M. Conventional and Explicitly Correlated ab Initio Benchmark Study on Water Clusters: Revision of the BEGDB and WATER27 Data Sets. *J. Chem. Theory Comput.* **2017**, *13*, 3136–3152.
- [194] Valdes, H. et al. Benchmark database on isolated small peptides containing an aromatic side chain: comparison between wave function and density functional theory methods and empirical force field. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2747.
- [195] Hostaš, J.; Řezáč, J. Accurate DFT-D3 Calculations in a Small Basis Set. *J. Chem. Theory Comput.* **2017**, *13*, 3575–3585.
- [196] Sedlak, R.; Janowski, T.; Pitoák, M.; Řezáč, J.; Pulay, P.; Hobza, P. Accuracy of quantum chemical methods for large noncovalent complexes. *J. Chem. Theory Comput.* **2013**, *9*, 3364–3374.
- [197] Huang, M.; Giese, T. J.; Lee, T. S.; York, D. M. Improvement of DNA and RNA sugar pucker profiles from semiempirical quantum methods. *J. Chem. Theory Comput.* **2014**, *10*, 1538–1545.
- [198] Gruzman, D.; Karton, A.; Martin, J. M. L. Performance of Ab Initio and Density Functional Methods for Conformational Equilibria of C_nH_{2n+2} Alkane Isomers (n = 48). *J. Phys. Chem. A* **2009**, *113*, 11974–11983.
- [199] Tsuzuki, S.; Honda, K.; Uchimaru, T.; Mikami, M. Estimated MP2 and CCSD(T) interaction energies of n-alkane dimers at the basis set limit: Comparison of the methods of Helgaker et al. and Feller. *J. Chem. Phys.* **2006**, *124*.
- [200] Goerigk, L.; Grimme, S. A general database for main group thermochemistry, kinetics, and noncovalent interactions - Assessment of common and reparameterized (meta-)GGA density functionals. *J. Chem. Theory Comput.* **2010**, *6*, 107–126.
- [201] French, A. D.; Johnson, G. P.; Stortz, C. a. Evaluation of Density Functionals and Basis Sets for Ga. *Society* **2009**, 679–692.
- [202] Řeha, D.; Valdés, H.; Vondrášek, J.; Hobza, P.; Abu-Riziq, A.; Crews, B.; De Vries, M. S. Structure and IR spectrum of phenylalanyl-glycyl-glycine tripeptide in the gas-phase: IR/UV experiments, Ab initio quantum chemical calculations, and molecular dynamic simulations. *Chem. - A Eur. J.* **2005**, *11*, 6803–6817.
- [203] Bryantsev, V. S.; Diallo, M. S.; Van Duin, A. C.; Goddard, W. A. Evaluation of B3LYP, X3LYP, and M06-Class density functionals for predicting the binding energies of neutral, protonated, and deprotonated water clusters. *J. Chem. Theory Comput.* **2009**, *5*, 1016–1026.

- [204] Rosskyt, P. J.; Karplus, M. Solvation. A Molecular Dynamics Study of A Dipeptide In Water. *J. Am. Chem. Soc.* **1979**, *101*, 1913–1937.
- [205] De M Seabra, G.; Walker, R. C.; Roitberg, A. E. Are current semiempirical methods better than force fields? A study from the thermodynamics perspective. *J. Phys. Chem. A* **2009**, *113*, 11938–11948.
- [206] Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; González-Outeiriño, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J. GLYCAM06: A generalizable biomolecular force field. Carbohydrates. *J. Comput. Chem.* **2008**, *29*, 622–655.
- [207] McNamara, J. P.; Muslim, A.-M.; Abdel-Aal, H.; Wang, H.; Mohr, M.; Hillier, I. H.; Bryce, R. A. Towards a quantum mechanical force field for carbohydrates: a reparametrized semi-empirical MO approach. *Chem. Phys. Letters* **2004**, *394*, 429–436.
- [208] Govender, K.; Gao, J.; Naidoo, K. J. AM1/d-CB1: A semiempirical model for QM/MM simulations of chemical glycobiology systems. *J. Chem. Theory Comput.* **2014**, *10*, 4694–4707.
- [209] Tsuneda, T.; Hirao, K. Long-range correction for density functional theory. *WIREs Comput. Mol. Sci.* **2014**, *4*, 375–390.
- [210] Csonka, G. I.; Perdew, J. P.; Ruzsinszky, A. Global hybrid functionals: A look at the engine under the hood. *J. Chem. Theory Comput.* **2010**, *6*, 3688–3703.
- [211] Kronik, L.; Kümmel, S. *Topics in Current Chemistry*; Springer, Berlin, Heidelberg, 2014; Vol. 347; pp 137–192.
- [212] Baer, R.; Livshits, E.; Salzner, U. Tuned Range-Separated Hybrids in Density Functional Theory. *Annu. Rev. Phys. Chem.* **2010**, *61*, 85–109.
- [213] Iikura, H.; Tsuneda, T.; Yanai, T.; Hirao, K. A long-range correction scheme for generalized-gradient-approximation exchange functionals. *J. Chem. Phys.* **2001**, *115*, 3540–3544.
- [214] Yanai, T.; Tew, D. P.; Handy, N. C. A new hybrid exchange-correlation functional using the Coulomb-attenuating method (CAM-B3LYP). *Chem. Phys. Letters* **2004**, *393*, 51–57.
- [215] Leininger, T.; Stoll, H.; Werner, H.-J.; Savin, A. Combining long-range configuration interaction with short-range density functionals. *Chem. Phys. Letters* **1997**, *275*, 151–160.
- [216] Savin, A. On degeneracy, near-degeneracy and density functional theory. *Theor. Comput. Chem.* **1996**, *4*, 327–357.
- [217] Gill, P. M. W.; Adamson, R. D.; Pople, J. A. Coulomb-attenuated exchange energy density functionals. *Mol. Phys.* **1996**, *88*, 1005–1009.
- [218] Niehaus, T. A.; Della Sala, F. Range separated functionals in the density functional based tight-binding method: Formalism. *Phys. Status Solidi Basic Res.* **2012**, *249*, 237–244.
- [219] Lutsker, V.; Aradi, B.; Niehaus, T. A. Implementation and benchmark of a long-range corrected functional in the density functional based tight-binding method. *J. Chem.*

- Phys.* **2015**, *143*, 184107.
- [220] Smith, D. G. A.; Burns, L. A.; Patkowski, K.; Sherrill, C. D. Revised Damping Parameters for the D3 Dispersion Correction to Density Functional Theory. *J. Phys. Chem. Letters* **2016**, *7*, 2197–2203.

Abbreviations

AO	atomic orbital	HREX	Hamiltonian Replica Exchange
ATP	adenosine triphosphate	LC	long-range corrected
BJ	Becke-Johnson damping	LCAO	linear combination of atomic orbitals
bR	bacteriorhodopsin ground state	LDA	local density approximation
CHARMM	Chemistry at Harvard Macromolecular Mechanics	MAD	mean absolute deviation
CPR	conjugate peak refinement	MD	molecular dynamics
dA	deoxyadenosine	MEP	Minimum energy path
dC	deoxycytidine	MM	molecular mechanics
DFT	density functional theory	MO	molecular orbital
DFTB	density functional tight-binding	MW	Multiple-Walkers
dG	deoxyguanosine	PES	potential energy surface
DNA	deoxyribonucleic acid	PMF	Potential of Mean Force
DP-DFTB	Delta-Pauli DFTB	POPC	1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine
dT	deoxythymidine	PRG	proton release group
GGA	generalized gradient approximation	PSO	particle swarm optimization
Gromacs	Groningen Machine for Chemical Simulations	PT	proton transfer
HBN	hydrogen-bonded network	rA	adenosine
HF	Hartree-Fock	rC	cytidine
		rG	guanosine

RMSD	root-mean-square deviation
RNA	ribonucleic acid
rU	uridine
SCC	self-consistent charge
QM	quantum mechanical
QM/MM	quantum mechanics/molecular mechanics
TIP3P	transferable intermolecular potential 3P
TS	transition state

Publications and Copyright

List of Publications

1. Van Vuong, Q.; Akkarapattiakal Kuriappan, J.; Kubillus, M.; Kranz, J.; Mast, T.; Niehaus, T. A.; Irle, S.; Elstner, M. “Parametrization and Benchmark of Long-Range Corrected DFTB2 for Organic Molecules.” *J. Chem. Theory Comput.*, **2017**, 14, 115–125

Copyright

Chapter 9 reproduced in part with permission from Van Vuong, Q.; Akkarapattiakal Kuriappan, J.; Kubillus, M.; Kranz, J.; Mast, T.; Niehaus, T. A.; Irle, S.; Elstner, M. “Parametrization and Benchmark of Long-Range Corrected DFTB2 for Organic Molecules.” *J. Chem. Theory Comput.*, **2017**, 14, 115–125. ©2017 American Chemical Society

Acknowledgements

At this point, I would like to take the opportunity and thank everyone who contributed to the successful completion of this work.

First of all, I am very grateful to Prof. Dr. Marcus Elstner for giving me the opportunity to engage in this field of research and for his supervision over the last years. He was always available for discussions, offered scientific advice and gave me the freedom to expand my research according to my own interest.

Special thanks goes to Dr. Tomáš Kubař for his continuous scientific support, all the fruitful discussions and the motivating words.

I would also like to express my gratitude to Prof. Dr. Qiang Cui (University of Wisconsin, Madison, USA) for the possibility to spend six months in his group. I really profited from Prof. Dr. Cui's expertise in QM/MM simulations and proton transfer reactions.

Furthermore, I have to thank all my colleagues in Karlsruhe and Madison for the nice atmosphere in the lab and not to forget out of the lab. It was always fun working and spending time together with you.

I thank Maximilian Kubillus for the helpful discussions and suggestions in the course of the Delta-Pauli parametrization.

I am also thankful to Sabine Holthoff for the great support with all kinds of administrative issues.

I am grateful to the Karlsruhe House of Young Scientists (KHYS) for providing me with a funding for my research stay abroad in Madison, USA.

Many computations were carried out on the JUSTUS High Performance Computing Facility located at the University of Ulm. With this, I would like to thank for the stable, perfectly maintained computer system and the provided computer time.

And finally, I would like to thank my girlfriend, my parents and my three sisters. That the family is mentioned at the end does not seem to reflect its true importance. In this spirit, thanks for your continuous support, patience and all the motivating words over the last years.