

# Spectroscopic Properties of Light-Activated Proteins Resolved by Multiscale-Simulations

Zur Erlangung des akademischen Grades einer  
DOKTORIN DER NATURWISSENSCHAFTEN

(Dr. rer. nat.)

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

genehmigte

DISSERTATION

von

M.Sc. Beatrix Mirinda Bold

1.Referent: Prof. Dr. Marcus Elstner

2.Referent: PD Dr. Florian Weigend

Tag der mündlichen Prüfung: 6. Mai 2019



---

## **Erklärung zur Dissertation**

Ich erkläre hiermit, dass ich die vorliegende Dissertation selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe. Weiterhin versichere ich, dass ich die Satzung des Karlsruher Instituts für Technologie (KIT) zur Sicherung guter wissenschaftlicher Praxis beachtet habe.

Karlsruhe, den 27.03.2019

Beatrix Mirinda Bold



---

## Acknowledgment

First of all, I would like to thank Prof. Dr. Marcus Elstner for giving me the opportunity to do my PhD thesis in his research group. I am thankful for all the interesting projects, in which I was involved. I am very grateful for all the inspiring discussions and hints that helped me in my work.

A big thank goes also to Dr. Tomáš Kubař and Dr. Sebastian Höfener for all the fruitful discussions and answering all my questions patiently. I would also like to thank Dr. Yanan Guo and Dr. Franziska E. Wolff, our Channelrhodopsin group, thanks for all the interesting discussions. Thanks a lot also to Dr. Marius Wanko for supportive advice. I express a special thanks to our light-harvesting group, Philipp Dohmen and Monja Sokolov, as well as to our cooperation group at the Jacob's university in Bremen. I am thankful for all the interesting discussions with Prof. Dr. Ulrich Kleinekathöfer and with Sayan Maity. I express also a special thanks to Sabine Holthoff, helping me with all administrative matters.

A lot of thanks to our previous "Girls Tower", Dr. Violetta Schneider, Dr. Franziska E. Wolff and Mila Andreeva. Thanks for all the fun we had the last years, the interesting discussions and of course the motivating "ice cream" breaks. I would like also to thank the whole TCB Team for the nice working atmosphere.

Vielen Dank an meiner ganzen Familie, meinen Eltern, meinem Bruder und meiner Oma für all das Verständnis und die Unterstützung während der gesamten Promotion. Ich möchte hier auch Frau Astrid Tari Fundis danken, die leider nicht mehr unter uns weilt, für ihre motivierenden Worte während des Studiums und am Anfang meiner Promotion.

Last but not least, I would like to thank all my colleagues for proof reading.



---

## Zusammenfassung

Rhodopsine und Lichtsammelkomplexe sind durch Licht aktivierte Proteine, die an natürlichen Prozessen wie dem Sehen oder der Photosynthese beteiligt sind. Um deren detaillierten Mechanismus zu verstehen, sind aufgrund ihrer Größe und der komplexen Umgebung, in die das Chromophor eingebettet ist, Multiskalen-Methoden erforderlich. Neue und effiziente Methoden sind daher sehr wünschenswert. In dieser Arbeit wurde somit die kürzlich erfolgte zeitabhängige Erweiterung der effizienten LC-DFTB Methode<sup>115</sup> für ihre Anwendung auf Rhodopsine und Lichtsammelkomplexe über einen systematischen Benchmark validiert. Zusätzlich wurden etablierte Methoden zur Untersuchung des *pre-gating* Prozesses des lichtgesteuerten Ionenkanals *Channelrhodopsin-2* Wildtyp (ChR2-WT) verwendet.

Die Benchmark Studie wurde sowohl an den isolierten Chromophoren Retinal und Bakteriochlorophyll a (BChl a), als auch am biologischen System als Ganzes durchgeführt. Dabei wurden Einflussfaktoren, die für ihre charakteristischen Absorptionsspektren verantwortlich sind, sowie Methoden berücksichtigt, die von *ab-initio* bis zum semiempirischen Ansatz reichen. LC-DFTB konnte qualitativ den gleichen Trend reproduzieren wie die LC-DFT Funktionale hinsichtlich der Beschreibung der Einflussfaktoren auf die Absorptionsspektren von Retinal und BChl a. Während LC-DFTB jedoch Schwächen in der Anwendung an Rhodopsinen zeigte, waren die Ergebnisse für Lichtsammelkomplexe in guter Übereinstimmung mit denen von *ab-initio* Methoden. LC-DFTB ist damit vielversprechend, mit dem Potenzial für weitere Untersuchungen.

Der *pre-gating* Prozess von ChR2-WT wurde mit Methoden untersucht, die sowohl auf einem klassischen als auch auf einem QM/MM Ansatz basieren. Somit ergaben sich aus QM/MM und klassischen MD Simulationen mehrere Teilzustände des ersten Zwischenzustands des Photozyklus. Nachfolgend, wurden mit diesen Strukturen UV/Vis- und Vibrationspektren berechnet, die in guter Übereinstimmung mit dem Experiment waren. Somit konnte ein möglicher *pre-gating* Prozess abgeleitet werden. Dies ermöglicht weitere Untersuchungen der folgenden Zwischenzustände des Photozyklus, um den detaillierten Mechanismus und die Funktion von ChR2-WT zu verstehen.



---

## Abstract

Rhodopsins and light-harvesting (LH) complexes are light-activated proteins involved in natural processes like vision or photosynthesis. To understand their detailed mechanism, multi-scale methods are needed due to their size and the complex environment in which the chromophore is embedded. New and efficient methods are thus highly desirable. In this work, the recent time-dependent extension of the efficient LC-DFTB method<sup>115</sup> was therefore validated for its application to rhodopsins and LH complexes via a systematic benchmark. Additionally, established methods were used for the investigation of the pre-gating process of the light-gated ion channel channelrhodopsin-2 wild-type (ChR2-WT).

The benchmark study was performed on both the isolated chromophores retinal and bacteriochlorophyll a (BChl a) as well as on the biological system as a whole. Influencing factors responsible for their characteristic absorption spectra as well as methods ranging from *ab-initio* to semiempiric levels of theory were taken into account. LC-DFTB could qualitatively reproduce the same trend as the LC-DFT functionals with regard to describe the influencing factors on the absorption spectra of retinal and BChl a. Even though, LC-DFTB showed weaknesses in the performance of rhodopsins, for LH complexes the results were in good agreement to those of *ab-initio* methods. LC-DFTB is therefore promising with the potential for further investigations.

The pre-gating process of ChR2-WT was investigated using methods based on both a classical as well as a QM/MM approach. Hence, QM/MM and classical MD simulations revealed several substates which are present in the first intermediate state of the photocycle. Subsequently, UV/Vis and vibration spectra were computed with these structures. The results were in good agreement with the experiment, suggesting a possible pre-gating process. This enables further investigations of the following intermediate states of the photocycle to understand the detailed mechanism and function of ChR2-WT.



---

# Contents

Zusammenfassung . . . . .	v
Abstract . . . . .	vii
<b>1 Introduction</b>	<b>1</b>
<b>2 Light-Activated Processes</b>	<b>3</b>
2.1 Photophysical and Photochemical Processes . . . . .	6
2.1.1 Isomerization . . . . .	7
2.1.2 Exciton Coupling and Exciton Transfer . . . . .	7
2.2 Biological Systems . . . . .	8
2.2.1 Excitation of Retinal and Bacteriochlorophyll . . . . .	8
2.2.2 Rhodopsins . . . . .	10
2.2.3 Light-Harvesting Complexes . . . . .	16
<b>3 Theoretical Background</b>	<b>21</b>
3.1 <i>ab-initio</i> Methods . . . . .	22
3.1.1 Hartree Fock Theory . . . . .	22
3.1.2 Configuration Interaction . . . . .	25
3.1.3 Coupled Cluster . . . . .	25
3.1.4 Multi-Reference Configuration Interaction . . . . .	26
3.1.5 Density Functional Theory . . . . .	27
3.1.6 Time-Dependent Density Functional Theory . . . . .	31
3.2 Semi-Empirical Methods . . . . .	33
3.2.1 Zerner's Intermediate Neglect of Differential Overlap . . . . .	33
3.2.2 Orthogonalization Model 2 . . . . .	34
3.2.3 Density-Functional Tight-Binding . . . . .	34
3.2.4 Time-Dependent Density Functional Tight Binding . . . . .	37
3.2.5 Long-Range Corrected Density Functional Tight Binding . . . . .	38
3.2.6 Time-Dependent Long-Range Corrected Density Functional Tight Binding . . . . .	39
3.3 Molecular Dynamics Simulation . . . . .	40
3.3.1 Molecular Mechanics . . . . .	40
3.3.2 Quantum Mechanics/Molecular Mechanics . . . . .	42

---

3.3.3	Geometry Optimization . . . . .	44
3.3.4	Equilibration . . . . .	44
3.3.5	Molecular Dynamics . . . . .	44
3.4	Spectroscopic Properties . . . . .	45
3.4.1	Excitation Energy . . . . .	45
3.4.2	Exciton Coupling . . . . .	46
3.4.3	Vibrational Frequencies . . . . .	49
<b>4</b>	<b>Benchmark of LC-DFTB on Retinal and Bacteriochlorophyll</b>	<b>51</b>
4.1	Introduction . . . . .	51
4.1.1	Computational Investigation of Color-Tuning . . . . .	52
4.1.2	Excitation Energy of Retinal . . . . .	53
4.1.3	Excitation Energy and Exciton Coupling of BChl a . . . . .	55
4.2	Computational Details . . . . .	59
4.2.1	Models and Geometry Optimization . . . . .	59
4.2.2	Excitation Energies . . . . .	59
4.2.3	Exciton Coupling . . . . .	60
4.3	Retinal . . . . .	62
4.3.1	Bond Length Alternation . . . . .	62
4.3.2	Twist of the $\beta$ -Ionone Ring . . . . .	63
4.3.3	Effect of the Electrostatic Environment . . . . .	65
4.4	Bacteriochlorophyll . . . . .	67
4.4.1	Bond Length Alternation . . . . .	67
4.4.2	Effect of the Electrostatic Environment . . . . .	69
4.4.3	Exciton Coupling . . . . .	72
4.5	Conclusion . . . . .	76
<b>5</b>	<b>Performance of LC-DFTB on Rhodopsins and Light-Harvesting Complex II</b>	<b>79</b>
5.1	Introduction . . . . .	79
5.2	Computational Details . . . . .	80
5.2.1	Models . . . . .	80
5.2.2	QM/MM MD Simulation . . . . .	80
5.2.3	MM MD Simulation . . . . .	81
5.2.4	Excitation Energies . . . . .	82
5.2.5	Coulomb Coupling . . . . .	82
5.3	Rhodopsins . . . . .	83
5.3.1	QM/MM Optimized Models . . . . .	84

5.3.2	QM/MM MD Simulation . . . . .	87
5.4	Light-Harvesting Complex II . . . . .	89
5.4.1	QM/MM Optimized Model . . . . .	89
5.4.2	QM/MM and MM MD Simulation . . . . .	92
5.5	Conclusion . . . . .	99
<b>6</b>	<b>The First Intermediate State of Channelrhodopsin-2</b>	
	<b>Wild-Type</b>	<b>101</b>
6.1	Introduction . . . . .	101
6.1.1	The Dark State . . . . .	102
6.1.2	Isomerization of all- <i>trans</i> Retinal . . . . .	104
6.1.3	Experimental Characterization of the First Intermediate State . . . . .	104
6.2	Computational Details . . . . .	106
6.2.1	Preparation of the P <sub>1</sub> State . . . . .	106
6.2.2	QM/MM MD Simulation . . . . .	107
6.2.3	MM MD Simulation . . . . .	109
6.2.4	Analysis of the Trajectories . . . . .	109
6.2.5	Excitation Energies . . . . .	110
6.2.6	Vibrational Frequencies . . . . .	110
6.3	QM/MM MD Simulations of the Early Substate P <sub>1a</sub> . . . . .	111
6.3.1	The Retinal Binding Pocket . . . . .	111
6.4	MM MD Simulations of P <sub>1</sub> . . . . .	121
6.4.1	Pre-Gating Process . . . . .	121
6.4.2	The Retinal Binding Pocket . . . . .	123
6.4.3	Helix Movements and Water Influx . . . . .	128
6.5	QM/MM MD Simulations of the Late Substate P <sub>1b</sub> . . . . .	130
6.5.1	The Retinal Binding Pocket . . . . .	131
6.6	Absorption Spectra of P <sub>1</sub> . . . . .	133
6.7	Characterization of P <sub>1a</sub> and P <sub>1b</sub> : Vibrational Frequencies . . . . .	135
6.8	Conclusion . . . . .	138
<b>7</b>	<b>Conclusion and Outlook</b>	<b>141</b>
<b>A</b>	<b>Appendix</b>	<b>143</b>
A.1	Benchmark of LC-DFTB on Retinal and Bacteriochlorophyll . . . . .	143
A.1.1	Retinal . . . . .	143
A.1.2	Bacteriochlorophyll . . . . .	144
A.2	Performance of LC-DFTB on Rhodopsins and Light-Harvesting Complex II . . . . .	149
A.2.1	Rhodopsins . . . . .	149
A.2.2	Light-Harvesting Complex II . . . . .	150

A.3 The First Intermediate State of Channelrhodopsin-2 Wild-Type . . . . .	151
A.3.1 QM/MM MD Simulations of the Early Substate P <sub>1a</sub> . . . . .	151
A.3.2 MM MD Simulations of P <sub>1</sub> . . . . .	153
List of Abbreviations . . . . .	157
List of Figures . . . . .	158
List of Tables . . . . .	161
<b>Bibliography</b>	<b>165</b>

---

## Introduction

Light is one of the most important natural energy source making life on earth possible. Photosynthesis of plants is one of the basic process which control climate. Humans, as well as most animals use light for vision and thus for orientation.

In chemistry, processes involving light are described as photochemical or photophysical processes. Deep into the 20th century, thermochemistry was playing a more dominant role than photochemistry.<sup>231</sup> However, during the 20th century quantum chemistry was becoming popular, allowing thus the description of chemical processes induced by electromagnetic radiation. While heat increases the microscopic motion of all molecules in a sample by activating their translational, rotational and vibrational modes, the absorption of light in the visible and UV spectrum is much more specific and has a significant effect on the reactivity of the absorbing molecule.<sup>176</sup>

Understanding photochemical or photophysical processes in biological systems becomes of great interest, since this knowledge helps engineers to adapt and modify these systems for applications in medicine or material science. In the beginning of the year 2000 a light induced cation channel called Channelrhodopsin from the unicellular green algae *Chlamydomonas reinhardtii* was found laying the basis for the field of optogenetic. Optogenetic makes the understanding of neural activity in brains and light therapies for medical purposes possible.<sup>155,156</sup> A deep understanding of photosynthesis processes in bacteria or plants by means of light-harvesting (LH) complexes provides new approaches for the design of efficient optoelectronic components or sensitizers for solar cells.

The study of biological systems involving light-activated processes is very complex. The light-absorbing part of the system, called chromophore is embedded in a complex protein environment. The size of the biological system and the different time scales on which biochemical processes occur are together challenging for both experiment and computation.

With the study of photochemical and photophysical processes many experimental spectroscopy methods were developed. One example is the UV/Vis spectroscopy which provide a way of analyzing the interaction between electromagnetic radiation and matter.<sup>4,231</sup> From ex-

---

periments, spectra or protein structures can be obtained. However, the detailed understanding of the spectra is not always given by experiments. Furthermore, the experimental conditions required for taking measurements can be detrimental to a protein's structure and function. For example, when proceeding at very low temperatures or when crystallization neglects the structurally or functionally relevant water environment due to the harsh conditions, drastic changes can occur in the protein fine structure.

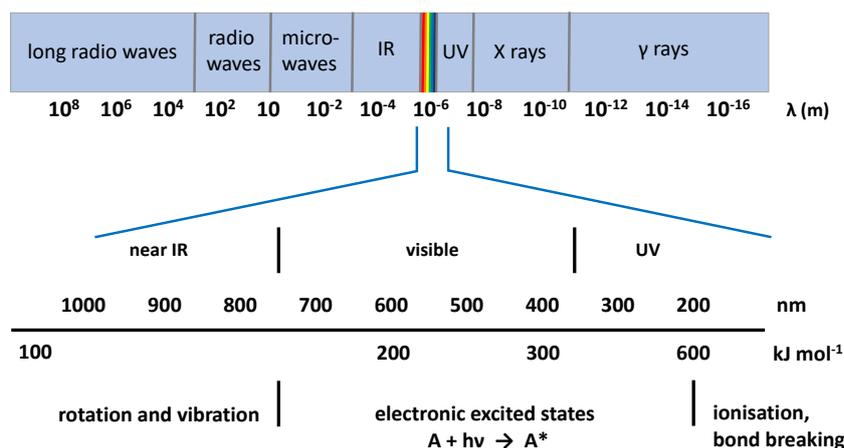
Therefore, computational chemistry became very important in this field. The development of different computational methods, differing in accuracy and computational costs, and the development of even faster computers provide the opportunity to simulate processes in biological systems on an atomistic scale and in their natural environment.

One part in this thesis concerns the benchmark of the recent time-dependent extension of the LC-DFTB method<sup>115</sup> (LC-TD-DFTB) for its application on the light-activated proteins rhodopsins and LH complexes. The other part deals with the investigation of the pre-gating mechanism of the photocycle of the light-gated cation channel channelrhodopsin-2 wild-type (ChR2-WT).

In chapter 2, an overview of the photochemical and photophysical processes occurring in biological systems is given and the two specific systems are introduced. Chapter 3 describes the theoretical background, including all computational methods used in this thesis. In chapter 4, a systematic benchmark of LC-TD-DFTB on retinal, the chromophore in rhodopsins and bacteriochlorophyll, the chromophore in LH complexes is described, taking several influencing factors into account responsible for shifting their absorption maxima. For this benchmark study various computational methods ranging from *ab-initio* to semiempiric levels of theory are used. In chapter 5, the benchmark study is extended in order to test the performance of LC-TD-DFTB in rhodopsins and LH complexes. Due to the low computational cost, LC-TD-DFTB is also able to compute excitation energies not only for a single structure but also for thousands of sampled structure. So that additionally molecular mechanical (MM) as well as quantum mechanical/molecular mechanical (QM/MM) molecular dynamic (MD) simulations are performed to account also for a statistical representation. Chapter 6 reports the investigation of the first intermediate state of ChR2-WT using several methods based on a MM as well as on a QM/MM approach. Furthermore, UV/Vis as well as vibration spectra are computed for the directly comparison to already existing experimental results. In the last chapter, chapter 7, a summary of this thesis as well as an outlook to possible future research questions is given.

## Light-Activated Processes

Photophysics and photochemistry describe the interaction of electromagnetic radiation with matter. While photophysics describes the mechanism of absorption and the behavior of molecular excited states, photochemistry concerns itself with the chemical reactions which follow from such an excitation. The electromagnetic spectrum can be classified in several regions dependent on the wavelength, see Fig. 2.1. In biological systems, chromophores absorb light belonging to the visible ( $\lambda \approx 400\text{-}700\text{ nm}$ ) and near UV region ( $\lambda \approx 200\text{-}400\text{ nm}$ ). The absorption of a photon transfers the molecule from the ground state to an excited state ( $A + h\nu \rightarrow A^*$ ). However, only those photons can be absorbed, whose energy are the same as the energetic difference between the two electronic states.<sup>103,176</sup>



**Figure 2.1** Spectrum of electromagnetic radiation by wavelengths. The near IR, visible and UV region is zoomed in. For this region, examples are given of different processes molecules undergo when activated by certain wavelengths.

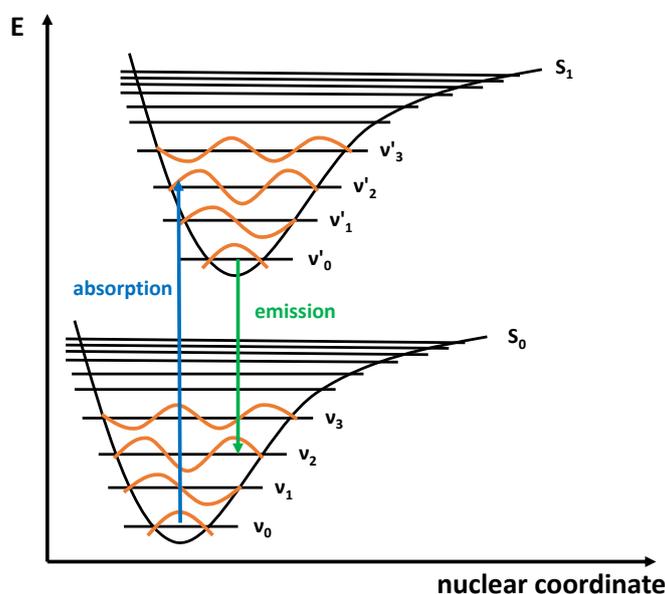
---

## The Franck-Condon Principle

The ground state as well as the excited states are described as potential energy surfaces (PES). Using quantum mechanics, the electrons and nuclei of the molecule are described by a wave function:

$$\Xi(r, R) = \Psi(r, R)\Theta(R) \quad (2.1)$$

where  $\Xi(r, R)$  is the total wave function,  $\Psi(r, R)$  the electronic wave function and  $\Theta(R)$  the nuclear wave function. Usually, the Born-Oppenheimer (BO) approximation is used. It uses the fact, that the nuclei have a larger mass than electrons and thus move slower. Hence, the electronic and the nuclear part of the equation can be separated. Using the BO approximation, the PES can be constructed (Fig. 2.2), where the minimum is the equilibrium distance of the nuclei.



**Figure 2.2** The Franck-Condon principle is illustrated here with two PES. The ground state  $S_0$  and the first excited state  $S_1$  are shown with several vibrational states. The vertical absorption (blue) to any vibronic state of  $S_1$  (here  $\nu'_2$ ) is displayed together with the vertical emission (green) to any vibronic state of  $S_0$  (here  $\nu_2$ ).

The electronic transition is faster than the nuclei can change their position. Hence, a vertical excitation (absorption) occurs, which is known as the Franck-Condon principle, see Fig. 2.2. This excitation occurs into a higher vibronic state, since the nuclei are no longer in the potential minimum after the excitation. Different excitations into different vibronic states happens and every excitation can be found in the absorption spectrum as part of the vibronical fine structure. After excitation, the molecule relaxes radiationlessly into the potential energy

---

minimum followed by electronic transition to the electronic ground state (emission). In the emission process several vibronic states of the ground state can be reached. As with the absorption, this can be seen in the emission spectrum. In the ideal case, the absorption and emission spectra are mirror images of each other, with the emission spectrum occurring in the longer wavelength region.<sup>103,176</sup> As seen in Fig. 2.2, both ground as well as the excited state are denoted "S<sub>n</sub>" with n= 0, 1, 2, . . . , which stands for "singlet". Molecules which have all electron spins paired ( $\uparrow\downarrow$ ), have a total electronic spin quantum number  $S = 0$ . This results in a multiplicity  $M$  ( $M = 2S + 1$ ) of 1 (singlet state). If the molecule has unpaired electrons, then  $S = 1$  and  $M = 3$  resulting in a triplet state ("T").<sup>103,176,231</sup>

### Dipole and Transition Moments

Two opposite charges  $+q$  and  $-q$  separated by a distance  $r$  constitute an electric dipole. The dipole moment  $\mu$  is a vector pointing from the negative to the positive charge and is calculated as  $\mu = qr$ . It is measured in debye ( $1 \text{ D} \approx 3.336 \cdot 10^{-30} \text{ Cm}$ ).

The calculation of the expectation value of the dipole moment operator using the ground state wavefunction  $\Psi_0$  gives the dipole moment of a molecule in the ground state:

$$\mu = \langle \Xi_0 | \hat{M} | \Xi_0 \rangle, \quad \hat{M} = e \left( \sum r_i - \sum Z_j r_j \right) \quad (2.2)$$

where the vector  $r_i$  represents the positions of the electrons  $i$ ,  $r_j$  the positions of the nuclei  $j$  and  $Z_j$  are the atomic numbers.

When a molecule absorbs a photon followed by a transition from the lowest vibrational level ( $\nu = 0$ ) of the electronic ground state  $n$  to any vibrational level ( $\nu'$ ) of the excited state  $m$ , the transition moment is given by  $M_{n,0 \rightarrow m,\nu'}$ :

$$M_{n,0 \rightarrow m,\nu'} = e \langle \Xi_{n,0} | \hat{M} | \Xi_{m,\nu'} \rangle \quad (2.3)$$

Taking the BO approximation into account, Eq. 2.3 is rewritten as:

$$M_{n,0 \rightarrow m,\nu'} = e \langle \Psi_{n,0} | \hat{M} | \Psi_{m,\nu'} \rangle \langle \Theta_0 | \Theta_{\nu'} \rangle \quad (2.4)$$

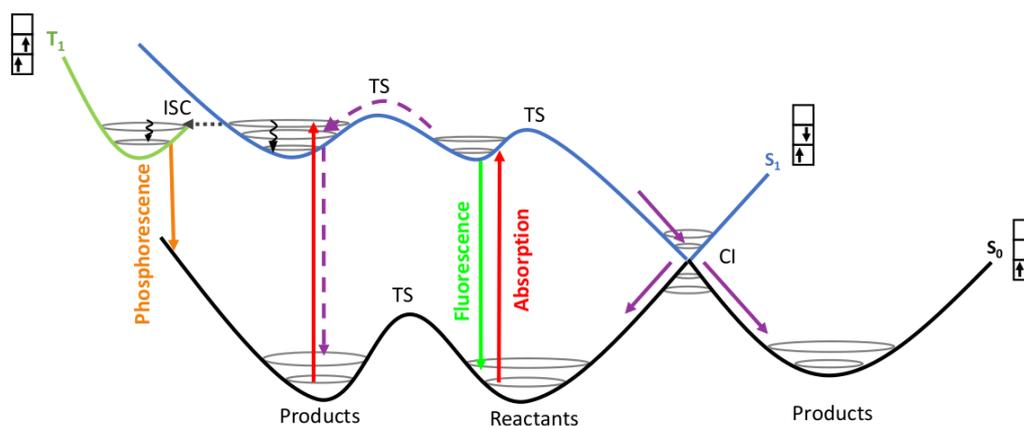
where  $\langle \Theta_0 | \Theta_{\nu'} \rangle$  is the overlap integral between the lowest vibrational level of the ground state and any vibrational level of the excited state, called Franck-Condon integral. Since several transitions into different vibronic states of the excited state are possible, the sum over  $\nu'$  of the Franck-Condon integral has to be taken into account. The oscillator strength  $f_{nm}$  is proportional to the square of the transition moment:

$$f_{nm} = \frac{8\pi^2 m_e \bar{\nu}_{nm}}{3hc^2} |M_{n \rightarrow m}|^2 \quad (2.5)$$

where the frequency  $\bar{\nu}_{nm}$  is the average value for the electronic transition. Strong transitions have an oscillator strength in the order of magnitude 1.<sup>75,103</sup>

## 2.1 Photophysical and Photochemical Processes

The different photophysical processes occurring after absorption of a photon are summarized in Fig. 2.3. In the following a brief overview is given. PES of the ground state as well as of the excited states are shown together with the vibrational states displayed as gray circles. Photophysical processes can be radiationless or radiative transitions. The radiationless transitions are shown as wavy lines, while the radiative ones are shown as solid lines. The isoenergetic transition is plotted as dashed black line. The processes involving the purple lines (solid and dashed) are explained in section 2.1.1.



**Figure 2.3** Schematic illustration of photochemical and photophysical processes.

After absorption into higher vibronic states of an electronic excited state, a so called vibrational relaxation will take place, where the molecule transfers the vibrational energy to the environment. The absorption is very fast (about  $10^{-15}$  s), while the vibration is slower and takes about  $10^{-11}$  to  $10^{-12}$  s. In the case of higher excited states than  $S_1$ , an internal conversion (IC) will take place ( $10^{-12}$  –  $10^{-6}$  s), i.e. a transition from one vibronic state to another vibronic state of the next lower-energy excited state followed by vibrational relaxation to the potential minimum, which is not shown in the figure. An isoenergetic transition between two states with different multiplicity is also possible, e.g. a transition from the vibrational ground state of  $S_1$  to an vibrationally excited state of  $T_1$  via spin reversal. This process is called intersystem crossing (ISC).<sup>103, 176, 231</sup>

Processes with radiative transitions are: (i) *Fluorescence* ( $10^{-9}$  –  $10^{-7}$  s): After absorption and vibrational relaxation into the first excited state minimum  $S_1$ , an emission into one of several vibronic states of the electronic ground state  $S_0$  can occur, as follows from the Franck-Condon principle. (ii) *Phosphorescence* ( $10^{-6}$  –  $10^{-3}$  s): After an ISC process from the vibrational ground state of  $S_1$  to a vibrationally excited state of  $T_1$  followed by vibrational relaxation, an emission from the triplet state  $T_1$  to the ground state  $S_0$  takes place.

### 2.1.1 Isomerization

The dashed purple arrows in Fig. 2.3 sketch the pathway of an adiabatic photochemical reaction, since the ground state PES and the excited state PES do not cross each other. After excitation, the entire reaction takes place on the excited state PES, as the system passes the transition state (TS) and the excited product is formed, which relaxes to its ground state radiatively or radiationlessly.

The solid purple arrows in Fig. 2.3 show a diabatic reaction, where ground and excited state PES are strongly coupled. Both PES can cross or nearly cross each other, also called conical intersection (CI). Therefore, the system relaxes radiationlessly back to the ground state  $S_0$  after excitation. Dependent on the energy barriers it can relax back to the reactant or to the product.<sup>103,231</sup>

The latter reaction pathway is an example for the *cis/trans* isomerization of a C=C double bond as found in rhodopsins, where the chromophore retinal isomerizes from 11-*cis* retinal to all-*trans* retinal in case of animal rhodopsins and from all-*trans* to 13-*cis* in case of microbial rhodopsins.<sup>202</sup>

### 2.1.2 Exciton Coupling and Exciton Transfer

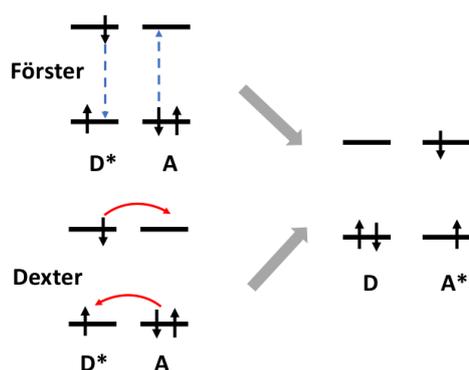
An excited molecule  $D^*$  (energy donor) can transfer its energy to a non-covalently bound neighboring molecule A (energy acceptor).



The so called Frenkel exciton describes a concept where the electron and the hole are located on the same molecule. Exciton coupling occur when two neighboring molecules energetically couple with each other. Thus, molecule A can become electronically excited without absorbing the incident light. In general, the process is isoenergetic, i.e. energy is not lost since it is transferred on the acceptor molecule A.

Interactions occurring between two neighboring molecules are exchange or Coulomb interactions. Energy transfer can happen as the so called Förster transfer or Dexter transfer, see Fig. 2.4. The former is found for long-range processes where dipole-dipole interactions are dominant. Hence, the energy is transferred via Coulomb interactions. The latter one is found for short-range processes where exchange interactions are dominant, i.e. the energy is transferred via electron exchange. The theoretical background about exciton couplings is given in chapter 3.

In biological processes, this kind of energy transfer is found for light-harvesting in photosynthetic organisms. In these organisms hundreds to thousands (bacterio)chlorophylls exist funnel the energy to a reaction center.<sup>75,103,145</sup>



**Figure 2.4** Two possible exciton transfer mechanisms, the Förster transfer and the Dexter transfer are shown, which can occur in organic molecules.

## 2.2 Biological Systems

### 2.2.1 Excitation of Retinal and Bacteriochlorophyll

In organic molecules several excitations are possible depending on the involved atom types, as listed in Tab. 2.1.

**Table 2.1** Possible excitations occurring in organic molecules.

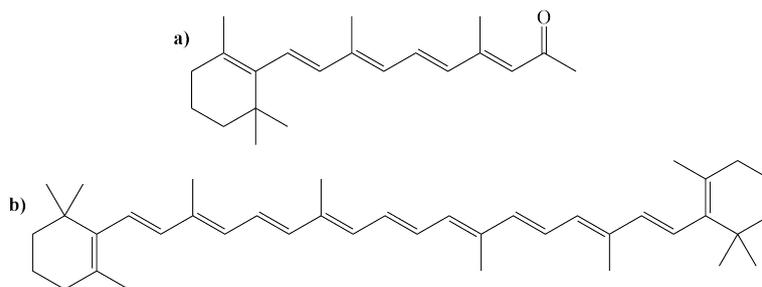
Excitation	Molecule type
$\pi \rightarrow \pi^*$	alkene, alkyne, aromatic compounds
$n \rightarrow \pi^*$	carbonyl group ( $-\text{C}=\text{O}$ ), thiocarbonyl group ( $-\text{C}=\text{S}$ ), imine group ( $-\text{C}=\text{N}$ )
$n \rightarrow \sigma^*$	amine, alcohol
$\sigma \rightarrow \sigma^*$	alkane

A characteristic absorption behavior is found for conjugated double bonds in linear polyenes or aromatic cyclic molecules. In case of linear polyenes the absorption maxima are shifted to longer wavelengths with increasing conjugation, i.e. when the chain length becomes longer and thus more double bonds are involved. For example, the absorption maximum of butadiene is found at a lower wavelength than the maximum of hexatriene. The increase of the conjugated chain narrows the energy gap between the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO). This in turn leads to absorption of longer wavelengths, since less energy is used for the excitation.

Linear polyenes can be equivalently depicted using the model of a particle in a box. The increase of the conjugated chain is accompanied by an increasing number of nodes. The  $\pi$ -electrons are delocalized over the complete chain. This delocalization of the  $\pi$ -electrons is also comparable with the metallic conductivity in solids, which leads to strong anisotropic

electrical polarization ability.<sup>75,103</sup>

The absorption maximum of ethylene, butadiene and hexatriene lie in the UV range. Above a certain chain length, the absorption maximum shifts to the visible range. This is found for chromophores in biological systems, e.g. retinal and  $\beta$ -carotene (Fig. 2.5), which both consist of long linear polyene chains. Both chromophores belong to the class of carotenoids, which absorb in the blue spectral region of about 420-570 nm leading to reddish, orange or yellow colors.<sup>41</sup>



**Figure 2.5** a) all-*trans* retinal and b)  $\beta$ -carotene.

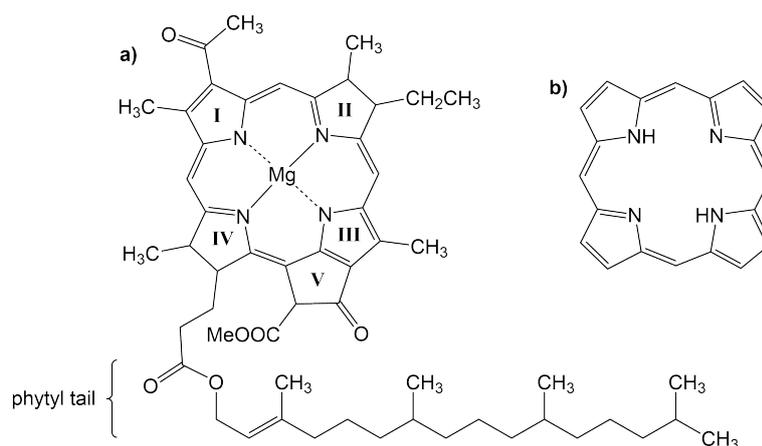
Aromatic cyclic molecules, like benzene, naphthalene, anthracene etc. also show an absorption maxima shift to longer wavelengths with increased  $\pi$ -electron systems, i.e. with an increasing size of aromatic rings. Nitrogen containing derivatives of benzene, naphthalene etc. exhibit  $n \rightarrow \pi^*$  excitations. This does not significantly affect the absorption bands compared to their nitrogenless equivalents.

The cyclic tetrapyrrole chromophore porphyrin (Fig. 4.2) is the basic structure of many important dyes and chromophores in nature, e.g. the chlorophylls (Chl) or bacteriochlorophylls (BChl) occurring in light-harvesting complexes. The (bacterio)chlorophylls contain additional a magnesium ion, see Fig. 4.2, which replace the two central hydrogen atoms. The absorption spectrum of porphyrins consist of the so called Q-band(s) in the visible region and a B-band (also called *Soret*-band) in the near UV region. This characteristic absorption is also found in (bacterio)chlorophylls. The labeling of the absorption peaks of (bacterio)chlorophylls is based on the four-orbital model for porphyrin derivatives as proposed by Gouterman *et al.*<sup>69,70</sup> In porphyrins, the two HOMOs (HOMO and HOMO-1) as well as the two LUMOs (LUMO and LUMO+1) are energetically degenerated, while in (bacterio)chlorophylls this degeneration is lifted due to a reduced symmetry. Hence the Q-bands are splitted up into a  $Q_x$  and  $Q_y$  band. Chlorophylls in general, absorb mainly in the blue or near-UV region and red or near-infrared spectral region leading to green colors.<sup>41,75,103</sup>

In Fig. 4.2 one BChl is shown, whose skeleton is the bacteriochlorin ring. Bacteriochlorin contains 18  $\pi$ -electrons and has a reduced aromaticity than porphyrin, which has 22  $\pi$ -electrons and whose aromatic system extends over the entire ring. Six different naturally occurring BChl (a, b, c, d, e, g) exist differing in their substituents. In this thesis, BChl a (Fig. 4.2)

is investigated, which contains a phytol tail, a polyisoprenoid alcohol chain esterified to the pyrrole ring IV.<sup>41,111</sup>

In general, the excitation of the chromophore from the ground state  $S_0$  to the first excited state  $S_1$  is accompanied by a change in bond orders and a change of the dipole moment. In the case of retinal, the excitation leads to a shift of the electron density along the polyene chain resulting in a high transition dipole moment of about 12 D.<sup>142</sup> In the case of BChl the ring shape reduces the changes in electron density leading to a transition dipole moment of 6.3 D.<sup>5,106</sup>



**Figure 2.6** a) Bacteriochlorophyll a (BChl a); b) Porphyrin.

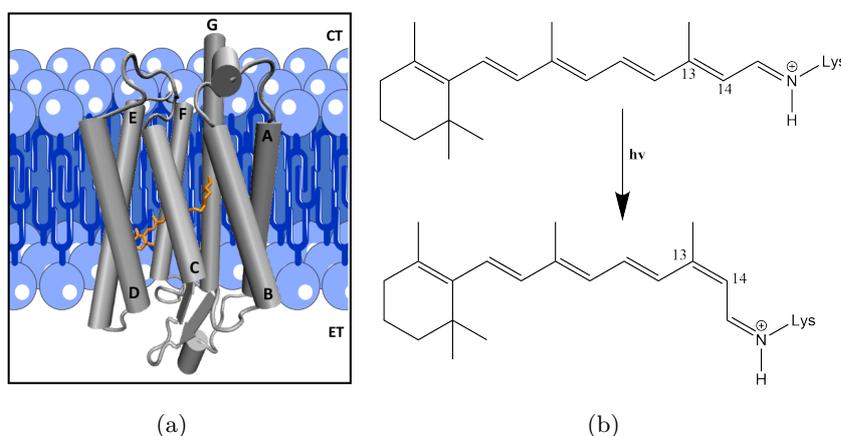
### 2.2.2 Rhodopsins

Rhodopsins, or also called retinylidene proteins are photoreceptor proteins involved in the sensing and response to light. The chromophore retinal absorbs photons for energy conversion or intra- and intercellular signaling. Rhodopsins are divided into two classes: (i) the microbial rhodopsins belonging to type I, found in archae, bacteria and fungi which function as light-driven ion pumps, light-gated ion channels or light sensors coupled to transducer proteins; and (ii) the animal rhodopsins belonging to type II found in animals and humans which are a subset of G-protein-coupled receptors. Examples for type I are bacteriorhodopsin (bR), a light-driven proton pump or halorhodopsin (hR), a chloride ion pump. Both are found in halobacteria, where also two rhodopsins important for photosensory signaling are found, sensory rhodopsin I (sRI) and sensory rhodopsin II (sRII), or also called pharaonis phoborhodopsin (ppR). The bovine rhodopsin (Rh) which are responsible for dim-light vision, or the human rod and cone visual pigments are examples for type II.<sup>54,202</sup>

All of these proteins are composed by seven transmembrane  $\alpha$ -helices embedded in a membrane and forming an internal binding pocket, where the chromophore retinal is bonded via a Schiff base linkage to the  $\epsilon$ -amino group of a lysine side chain of helix G, see Fig. 2.7(a). The chromophore retinal is usually protonated (RSBH<sup>+</sup>). Changes in its protonation state

are responsible for signaling or transport activity of the rhodopsins. The primary event in rhodopsins is the isomerization of retinal after photon absorption. For efficient light to energy or light to signal conversion, rhodopsins differ in the retinal isomerization pattern, retinal conformations, internal water arrangement and protein environment.<sup>54</sup> In microbial rhodopsins the retinal is isomerized from an all-*trans* to a 13-*cis* configuration (Fig. 2.7(b)), while in animal rhodopsins it is isomerized from 11-*cis* to all-*trans*.

In this thesis, only the first class of rhodopsins, belonging to the microbial rhodopsins are investigated. From this type I class, bR and ppR serve as models for the benchmark study of LC-TD-DFTB. A third rhodopsin, the light-gated cation channel channelrhodopsin-2, is investigated in chapter 6.



**Figure 2.7** (a) Schematic representation of a rhodopsin embedded in a membrane, retinal in orange is linked to helix G, CT means cytoplasm and ET means extracellular. (b) all-*trans* retinal bonded via a Schiff base linkage to the  $\epsilon$ -amino group of a lysine side chain is isomerized to 13-*cis* retinal around the  $C_{13}=C_{14}$  double bond after photon absorption.

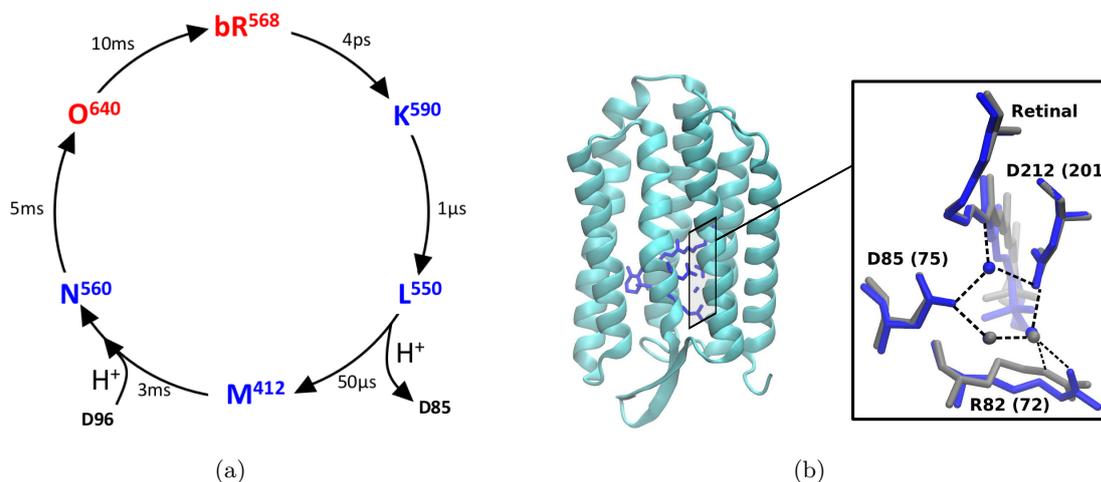
### Bacteriorhodopsin and Phoborhodopsin

Bacteriorhodopsin (bR) and pharaonis phoborhodopsin (ppR) are found in halobacterium, *Halobacterium salinarum*. The proton pump activity of bR lead to a formation of a membrane potential, which is in turn used for light-energy conversion, i.e. it is used for ATP synthesis. ppR works as negative phototaxis sensor.<sup>54</sup>

bR, which was found in the early 1970s, was the first discovered microbial rhodopsin,<sup>170</sup> whose structural arrangement was determined by electron microscopy<sup>79</sup> and the first one whose amino acid sequences was known.<sup>102</sup> Since then, it is the best understood and studied microbial rhodopsin and serves often as a reference model for the study of other rhodopsins.

The activation of bR and ppR by light triggers several reactions summarized in a so called photocycle, as shown for bR in Fig. 2.8(a).<sup>38, 121, 143</sup> Various experimental techniques, like X-ray crystallography, UV/Vis absorption, Fourier-transformed infrared (FTIR), Raman or nuclear

magnetic resonance (NMR) spectroscopy as well as various computational methods were used to determine and study different intermediate states of the photocycle.<sup>7, 82, 140, 143, 162, 228, 233</sup> Here a short summary of the events in the photocycle of bR is given. It has to be pointed out, that also several substates exist, which are spectroscopically silent but could be identified using computational studies.<sup>228, 233</sup>



**Figure 2.8** (a) Photocycle of bR, in red all-*trans* retinal and in blue 13-*cis* retinal; the numbers next to the arrows belong to the lifetime of the respective state and the superscript numbers belong to the absorption maxima. (b) Active site of bR (blue) and ppR (gray) in the dark state.

The primary event is the isomerization of retinal, followed by structural changes in the red-shifted K intermediate state which are mainly characterized by hydrogen bond changes or distance changes.<sup>27, 228</sup> The subsequent blue-shifted L intermediate is formed within 1 μs and is the precursor for the deprotonation of the RSBH<sup>+</sup> to reach the next strongly blue-shifted intermediate state M. The proton acceptor of the RSBH<sup>+</sup> is the amino acid D85.<sup>25, 68, 136</sup> The formation of the M state allows protons to pass the membrane vectorially from the cytoplasmatic side to the extracellular side. The next step in the photocycle is the reprotonation of the RSBH<sup>+</sup> by the proton donor D96 and the formation of the intermediate state N.<sup>33, 148</sup> Subsequently, the 13-*cis* retinal is isomerized back to the all-*trans* retinal leading to the red-shifted O state.<sup>201</sup> Finally, D96 is reprotonated from the cytoplasmatic medium. In the last step O→bR, a proton transfer from D85 to the proton-release complex takes place.<sup>121</sup> It is worthwhile to mention here, that in the dark state, i.e. the state without light contact, contains a 2:1 mixture of all-*trans*,15-*anti* and 13-*cis*,15-*syn* retinal, where the latter also undergoes a photocycle but has no influence on the pump activity of bR.<sup>54</sup>

Both, bR and ppR share the same basic structure. The so called active site includes the RSBH<sup>+</sup> and two counterions. A common feature of all microbial rhodopsins is the water presence in the active site. In the dark state of bR and ppR, the RSBH<sup>+</sup> forms a typical pentagonal

hydrogen bond network with three water molecules and two counterions,<sup>57,99,198,199,234</sup> see Fig. 2.8(b). Even though bR and ppR share the same basic structure, the absorption maxima of both differ by about 70 nm.<sup>38</sup> Details about structural differences in the binding pocket and the reason for the shifts are introduced in chapter 5.

## Channelrhodopsin-2

Channelrhodopsins (ChRs) are light-gated cation channels and firstly found in the unicellular green algae *Chlamydomonas reinhardtii*. ChRs are conductive for protons as well as monovalent and divalent cations ( $\text{Na}^+$ ,  $\text{K}^+$ ,  $\text{Ca}^{2+}$ ) resulting in plasma membrane depolarization.<sup>155,156</sup> They are located in the eyespot of the motile unicellular algae and serve there as a sensory photoreceptor allowing the algae to identify optimal light conditions for growth.<sup>189</sup> The unique combination of both photosensor and channel within a single protein paves the way for optogenetics. It is a research field using cell-specific promoters to express light-activatable proteins to be able to selectively activate or silence cells through the application of short light pulses.<sup>26,54,124,154</sup> In algae, two types of ChRs are found, termed channelrhodopsin-1 (ChR1) and channelrhodopsin-2 (ChR2). Even though ChR1 occurs more frequently, ChR2 is currently the most investigated ChR because of its superior expression in host cells.<sup>54</sup>

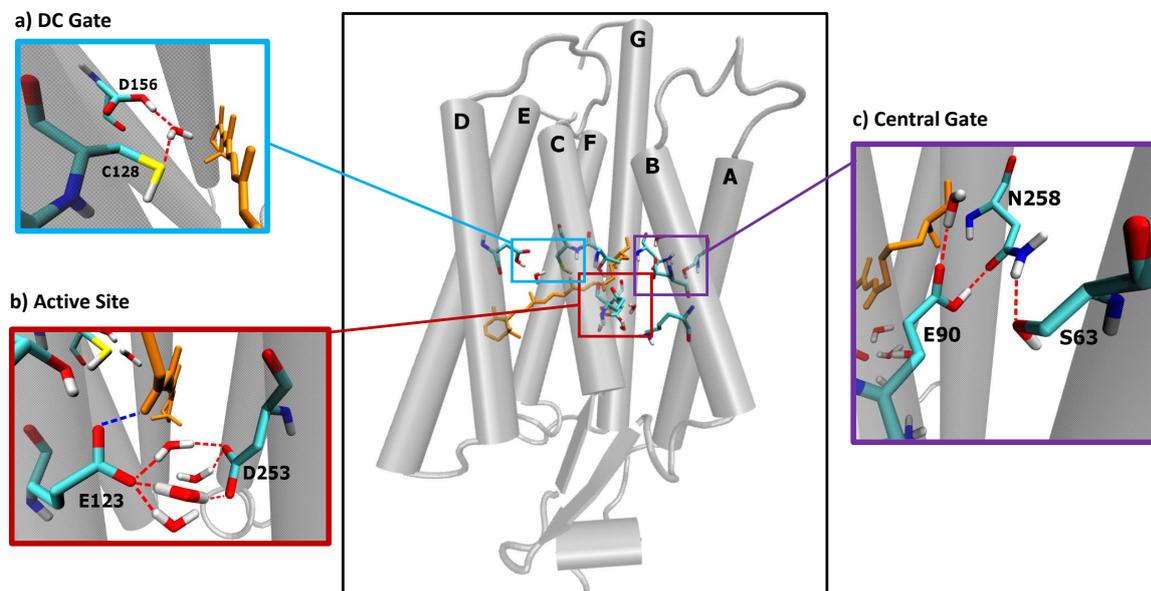
As typical for rhodospins, ChRs contain seven transmembrane  $\alpha$ -helices (A-G) embedded in a membrane forming an internal pocket, where the retinal chromophore is covalently bound to a lysine residue of the seventh helix (G) through a protonated Schiff base (RSBH<sup>+</sup>). Electron microscopy reveals that ChRs exhibit a dimeric structure, with the third and the fourth helices (C and D) placed at the dimer interface.<sup>151</sup>

In contrast to bR, no crystal structure of ChR2-WT was available until 2017.<sup>217</sup> An alternative is the hybrid model C1C2,<sup>100</sup> an X-ray structure comprised of the first 5 helices (A-E) of ChR1 and the last 2 helices (F-G) of ChR2. Another approximation is given by an homology model based on the C1C2 structure and obtained by sampling with classical force field MD simulations and structural refinement with combined QM/MM simulations.<sup>224</sup> Deeper structural comparison of the X-ray structure and the C1C2 as well as the homology model are given in chapter 6. The comparison of the bR crystal structure to the C1C2 structure reveals similarities for helices C-F, however differences of helices A and B which show a remarkably large number of glutamates (E82, E83, E90, E97, E101) in ChRs. Furthermore, helices A and B are tilted outward and helix G is shifted towards the central axis of the monomer compared to bR. This suggests, that helices A, B and G form a cavity, which allows water influx for a cation-translocation pathway.<sup>100</sup>

## Structural Motifs of ChR2-WT

The structure of ChR2 exhibits a characteristic active site in the dark state as well as some functional gates important for the channel functionality discovered by experimental studies of

mutants of ChRs.<sup>54,189</sup> Fig. 2.9 shows all important structural motifs.



**Figure 2.9** One monomer of ChR2-WT is shown as found in the homology model, zoomed in are: a) the water-bridged DC gate; b) active site, here hydrogen bond of RSBH<sup>+</sup> to E123 and c) central gate, here E90 orientated upward towards the cytoplasmic side.

The *active site* of ChR2 contains the RSBH<sup>+</sup>, two negatively charged counterions E123 (D85 in bR) and D253 (D212 in bR) and some water molecules (Fig. 2.9 (b)). The presence of water molecules was also highlighted by experimental studies.<sup>216</sup> A QM/MM study of the homology model on the active site structure in the dark state of ChR2 reveals a flexible hydrogen bond pattern of the RSBH<sup>+</sup>, i.e. hydrogen bonded to either E123, D253 or a water molecule.<sup>73</sup>

The *central gate* contains the residues S63, E90 and N258 forming a hydrogen bond network, as shown in Fig. 2.9 (c) with an upward orientation of E90, as found in the C1C2 X-ray structure and the homology model. The main difference observed in the ChR2 X-ray structure is the downward orientation of E90 toward E123, which is discussed in chapter 6. The E90Q mutant reveals changes in the ion selectivity displaying an increased selectivity for Na<sup>+</sup> as compared to the wild-type.<sup>71</sup>

The residues E82, E83, H134 and H265 form the inner gate, containing an additional cation binding site<sup>225</sup> and an interhelical hydrogen bond between the residues E83 and H134. Similarly to the central gate, the mutant H134R showed changes in ion selectivity.<sup>71</sup>

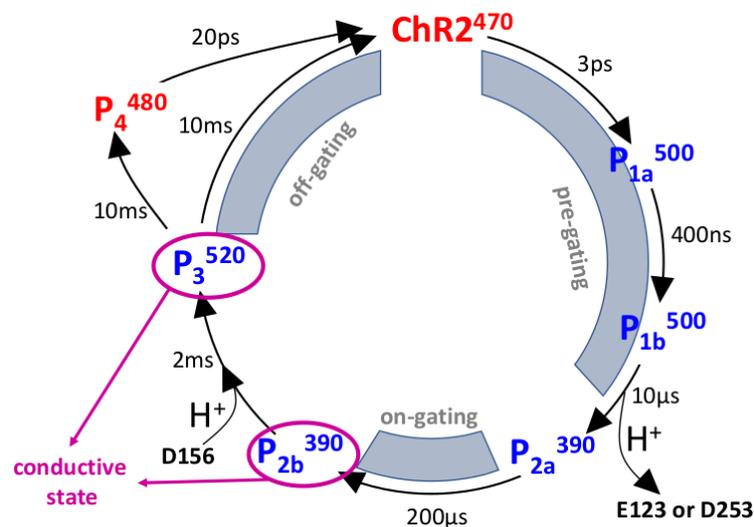
The *DC gate* contains the residues C128 and D156 bridged via a water molecule, see Fig. 2.9 (a) as obtained computationally by vibrational frequency analysis<sup>225,228</sup> and later confirmed in the ChR2 X-ray structure.<sup>217</sup> The importance of the DC Gate for on- and off-gating is reflected by studies on the C128T mutant.<sup>182,203</sup> These studies showed a 200- to

10,000-fold extended lifetime of the conducting state compared to ChR2 wild-type (ChR2-WT). In general, mutation of one or more residues on the one hand provides an insight into changes of channel activities and on the other hand opens the door for engineering of further novel and useful ChR properties. For example, the mutation of C128 leads cells to depolarize readily at quite low light intensities, which is useful for application in neuroscience.<sup>23</sup>

### Photocycle of ChR2-WT

The activation by light triggers several reactions involving the opening and closing of the channel similar to the other microbial rhodopsins, which are summarized in a photocycle. The different intermediate states are identified by UV/Vis spectroscopy as well as FTIR, Raman or NMR spectroscopy.<sup>20, 116, 132–134, 152, 153, 161, 180, 181, 216</sup> In contrast to bR, several open questions exist, since the intermediate states are mainly identified spectroscopically and only few computational studies exist based on the C1C2 structure until now.<sup>13, 37, 208</sup> Hence, only the main intermediate states are characterized. However, a detailed atomistic study is still missing.

Fig. 2.10 shows a photocycle of ChR2-WT as proposed by Lórenz-Fonfría et al.<sup>135</sup> using time-resolved UV/Vis, step-scan FTIR and tunable quantum cascade laser (QCL) IR spectroscopies. The primary event after photon absorption is the isomerization of all-*trans* retinal to 13-*cis* retinal. The photocycle consists of four different intermediate states,  $P_1^{500}$ ,  $P_2^{390}$ ,  $P_3^{520}$  and  $P_4^{480}$ . The first and the second intermediates are divided into two substates *a* and *b*. The kinetics of the photocycle is similar to those of bR.<sup>216</sup>



**Figure 2.10** Photocycle of ChR2-WT. The states in red contain an all-*trans* retinal configuration, while the ones in blue contain a 13-*cis* retinal configuration. The numbers next to the arrows belong to the lifetime of the respective state and the superscript numbers belong to the absorption maxima.

The formation of the red-shifted first intermediate state within 3 ps after retinal isomerization accompanies with the formation of the pre-gating process, which is preserved until the formation of the blue-shifted state  $P_2$ . The first intermediate state  $P_1$  might include some large structural rearrangement or distortion of helices, which is unusual for microbial rhodopsins. This however is suggested to be necessary for the formation of a larger pore for the conduction of cations.<sup>132,161,180,183</sup> Furthermore, the formation or enlargement of water-filled pores inside the protein might be a prerequisite for cation permeation.<sup>134</sup>  $P_1$  is also the precursor for the deprotonation of the  $RSBH^+$ , especially the late substate  $P_{1b}$  which might be the analog to the L state in bR. The primary proton acceptor is suggested to be D253 (D212 in bR) as resulted by FTIR spectroscopy in contrast to bR, where D85 serves as the proton acceptor.<sup>134</sup> Studies involving mutants of E123 (E123Q, E123T and E123A) observed, that E123 serves as the voltage sensor regulating the photocycle speed at different membrane voltages.<sup>72</sup> The pre-gating process paves the way for opening of the channel, which occurs with the formation of the first substate of  $P_2$  (M state in bR), beginning the on-gating process. The conductive states are allocated to the second substate of  $P_2$  and the red-shifted state  $P_3$ .<sup>17</sup> The reprotonation of the  $RSBH^+$  takes place between  $P_{2b}$  and  $P_3$ , where the proton donor is suggested to be D156, since its kinetics of deprotonation correlate with the kinetics of the reprotonation of RSB.<sup>134</sup> The channel closes after  $P_3$  followed by the recovery of all-*trans* retinal, which lead either directly to the dark state or to the formation of the intermediate state  $P_4$  only 10 nm red-shifted to the dark state.<sup>132</sup>

Analogously to bR, the dark state of the photocycle consists of a mixture of all-*trans*,15-*anti* and 13-*cis*,15-*syn* retinal in a 70:30 ratio. The photocycle originating from the 13-*cis*,15-*syn* retinal configuration is very fast, which supports the hypothesis that only the all-*trans*,15-*anti* photocycle is linked to the conductive activity of ChR2, as was also concluded for bR.<sup>132,150,152</sup>

### 2.2.3 Light-Harvesting Complexes

Plants, algae and bacteria use efficiently the energy of sunlight to convert it into chemical energy. This process is called photosynthesis. The functional unit of photosynthesis is build up of light-harvesting (LH) complexes, also called antenna complexes.

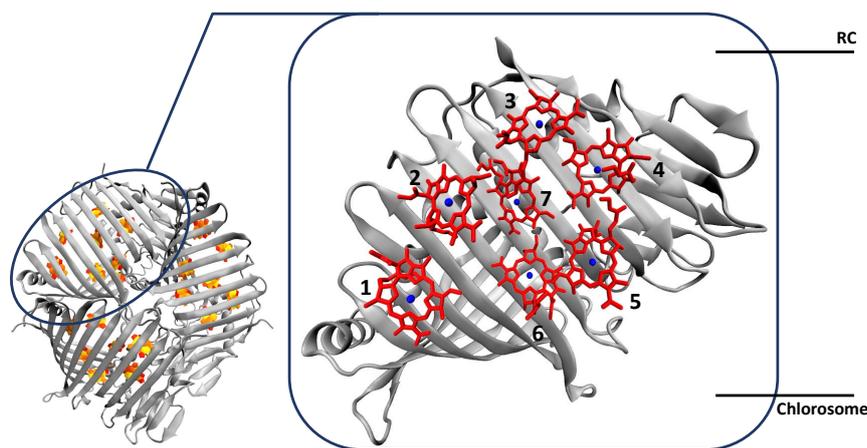
LH complexes absorb the energy of sunlight and transfer it efficiently in a range of picoseconds to a photosynthetic reaction center (RC). There, the energy is used for charge separation to initiate further electron transfer steps for the synthesis of chemical compounds.<sup>41</sup> LH complexes contain different chromophores to cover a broad range of the visible spectrum of light. These are: (bacterio)chlorophylls, carotenoids, and phycobilins. Additional, LH complexes consist of numerous chromophores, since only one chromophore could not provide such an efficient process. The number of chromophores varies from system to system in a range of 10 to 100.<sup>91</sup> Furthermore, the efficiency is also provided by several LH complexes together in one system, which are in turn differently arranged dependent on the organism.<sup>41,43,91,111,160</sup>

In this thesis, two LH complexes of bacteria are used, which are the Fenna-Matthews-Olsen Complex (FMO) of green sulfur bacteria and the light-harvesting complex II (LH2) of purple bacteria. LH complexes of bacteria perform an anoxygenic photosynthesis in contrast to plants, i.e. without oxygen production.<sup>41</sup> The FMO complex is used for the benchmark of LC-TD-DFTB and therefore only briefly introduced, while the LH2 complex is also subjected to a QM/MM study, see chapter 5.

### Fenna-Matthews-Olsen Complex

The FMO complex of green sulfur bacteria was the first LH complex for which an X-ray structure was available.<sup>55,144</sup> The green sulfur bacteria contains the largest naturally occurring light harvesting antenna system. It consists of chlorosomes, which in turn consist of hundreds to thousands of BChl chromophores. The FMO complexes connect the chlorosomes with the reaction center linked to one side of the membrane and therefore act as an energy-transfer unit.

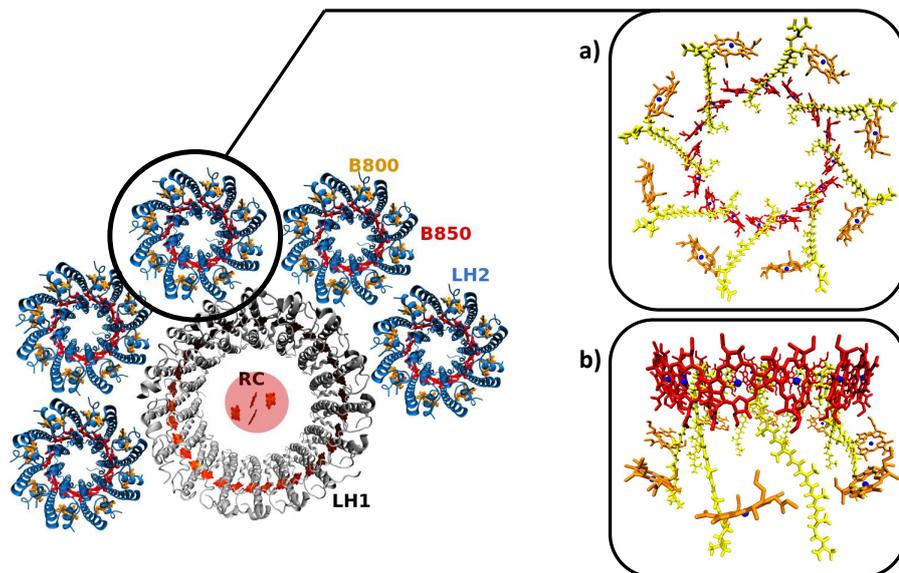
The FMO complex is set up as a trimer arranged in a  $C_3$  symmetry as shown in Fig. 2.11. Each monomer contains eight BChl a chromophores. The function of the eighth chromophore of the FMO complex is until today controversially discussed in literature, since it is further away from the other BChls.<sup>62,94,171,188</sup> Thus, it is not known if it plays a significant role in the energy transfer. However, it could serve as a linker molecule to the chlorosome.<sup>91</sup> The distances between the seven BChl a chromophores are in a range of about 10 to 20 Å. In an experimental study, the absorption range of the BChl a chromophores was determined to be about 790 to 825 nm (1.50-1.57 eV).<sup>210</sup>



**Figure 2.11** Schematic presentation of the FMO trimer. Here, one FMO monomer (PDB 3EOJ) of *Prosthecochloris aesturii* is zoomed in to visualize the arrangement of the seven BChl chromophores. The eighth BChl chromophore is not shown here, since it is further away, located behind chromophores 5 and 6.

## Light-Harvesting Complex II

The LH2 complex is embedded in a membrane of purple bacteria. Multiple LH2 complexes act as subunits surrounding the so called LH1 complex, see Fig. 2.12. Both LH complexes have a symmetric ring shape. The reaction center is linked to the LH1 complex. In some species, the LH1 complex has a C-shape or forms a S-shaped dimer surrounding the reaction center.



**Figure 2.12** Left: schematic presentation to visualize the arrangement of the LH1 (PDB 3WMM) and LH2 (PDB 1LGH) complexes in purple bacteria (top view). Right: one LH2 is zoomed in to display the chromophores: BChla, in red: B850 ring and orange: B800 ring; and carotenoides in yellow. a) top view; b) side view.

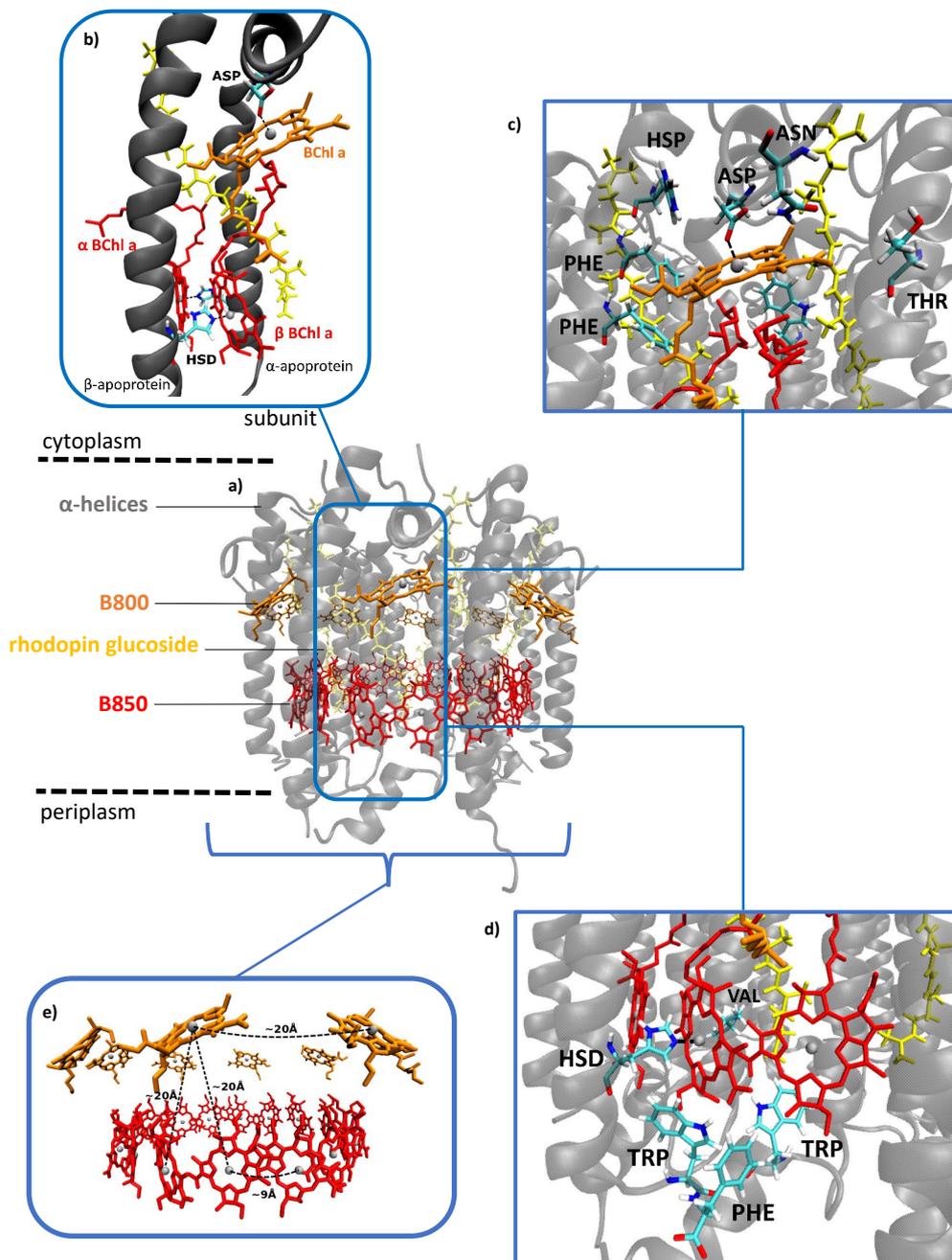
The first X-ray structures were resolved of the LH2 complexes of *Rhodospseudomonas (Rps.) acidophila*<sup>146</sup> and *Rhodospirillum (Rs.) molischianum*.<sup>109</sup> The latter is used for the study in this thesis. LH2 complexes contain several BChla and carotenoides (rhodopin glucoside), where the BChla chromophores are arranged in two rings. The two BChla rings in the LH2 complex exhibit characteristic absorption maxima at 800 nm and 850 nm and thus are referred to the B800 ring and B850 ring. In *Rs. molischianum*, the B800 ring contains 8 BChla chromophores, which are separated by a distance of about 20 Å, measured from Mg to Mg ion, visualized as the BChla ring in orange of Fig. 2.13 (b). This ring is oriented parallel to the membrane plane and perpendicular to the other BChla ring, the B850 ring. The B850 ring contains 16 BChla chromophores, which are more compactly arranged. The distance between them is about 9-10 Å, also measured from Mg to Mg ion and shown as the BChla ring in red of Fig. 2.13 (b).

The chromophores are arranged in so called  $\alpha - \beta$  subunits as shown in Fig. 2.13 (c) consisting of two  $\alpha$ -helices, three BChla chromophores and one carotenoid. Two BChla

chromophores originate from the B850 ring and the remaining one from the B800 ring. The BChl a chromophores are additionally distinguished depending on which  $\alpha$  or  $\beta$  apoprotein they are linked to. The ones linked to the  $\alpha$  apoprotein are denoted as  $\alpha$ -BChl a, pointing with their phytyl tail towards the inner ring. The other, which are linked to the  $\beta$  apoprotein are termed  $\beta$ -BChl a, where their phytyl tail is pointing towards the outside of the ring.

Depending on the species, the number of the chromophores and thus also the number of subunits in one LH2 complex differ. The LH2 complex of *Rs. molischianum* shows an 8-fold symmetry, while that of *Rps. acidophila* shows a 9-fold symmetry. The diameter of the LH2 complex is about 65 Å and also varies with species. The LH1 complex, in contrast, has a diameter of about 120 Å and about 16-18 subunits. The LH1 ring shows an absorption maximum of about 875 nm.

The LH2 complex serves as an example how excitonic interaction can tune an absorption spectrum,<sup>43</sup> which is referred to the B850 ring (1.46 eV). The absorption maximum of the B800 ring (1.55 eV) is set to be mainly influenced by the protein environment, since it is near to the BChl a monomer absorption maximum, which is about 770 nm (1.61 eV) in organic solvents.<sup>213</sup>

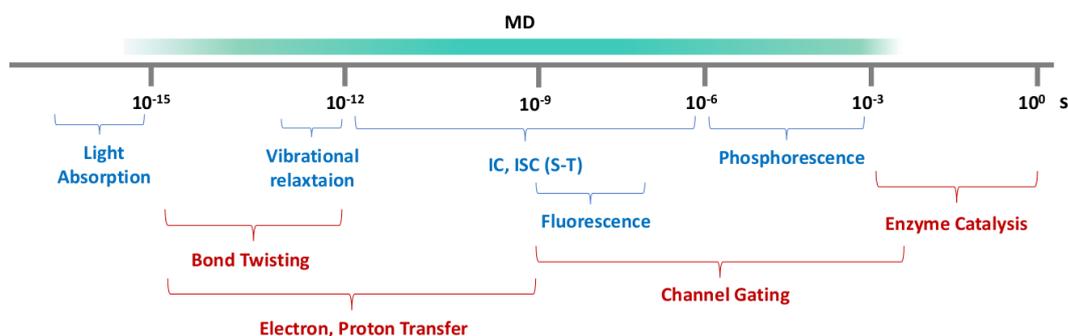


**Figure 2.13** Structural motifs of the LH2 complex: a) side view of the LH2 complex; b) subunit of the LH2 complex; c) One BChl<sub>a</sub> of the B800 ring zoomed in with its surrounded amino acids; d) Two BChl<sub>a</sub> of the B850 ring zoomed in with the surrounded amino acids; e) Distances of the BChl<sub>a</sub> chromophores of the B800 and B850 ring.

---

## Theoretical Background

Computational methods allow us to investigate biochemical processes on an atomistic scale. The investigation of biochemical processes requires multiscale methods, since they cover a broad range of time scales from femtoseconds for the absorption of light to seconds for enzyme catalysis, as shown here:



**Figure 3.1** Examples of biochemically relevant processes spanning a broad time scale. Photochemical and photophysical processes are shown in blue and other processes in red.

Photophysical and photochemical processes in biological systems require *ab-initio* methods to yield an accurate description of the excited state properties. Due to the high computational effort, *ab-initio* methods are limited to the number of atoms ( $\sim 100$ ). Hence, semi-empirical methods are needed, that can handle larger systems and allow the investigation of sampled structures. For the sampling of several structures, molecular dynamic (MD) simulations based on force fields have been established over the years in order to elucidate proteins in their structure and thermodynamics. One advantage of classical MD simulations is that especially large molecules can be described, as found in biological systems, since the computational effort is much lower than in *ab-initio* simulations. Thus, it is possible to simulate systems including up to 100 000 atoms for up to several milliseconds. In order to describe even larger

systems or longer time scales, so-called coarse grained methods are available. However, these methods were not employed in this work here. In the following, several methods ranging from *ab-initio* to semiempiric as well as MD simulations are presented, including the methods used in this thesis.

### 3.1 *ab-initio* Methods

#### Wavefunction Based Methods

In quantum mechanics, the Schrödinger equation describes how the quantum state of a non-relativistic system changes with time. If we describe a stationary state of the Hamiltonian the time-dependent Schrödinger equation can be transformed into the time-independent Schrödinger equation:

$$\hat{H}(r)\Psi(r) = E(r)\Psi(r) \quad \text{with} \quad \hat{H}(r) = \hat{T}(r) + \hat{V}(r) \quad (3.1)$$

The Hamilton operator  $\hat{H}$  for a many-electron system consists of the kinetic energy ( $\hat{T}$ ) as well as the potential energy ( $\hat{V}$ ) of  $N_e$  electrons ( $i, j$ ) and  $N_k$  nuclei ( $A, B$ ) and reads in atomic units:

$$\hat{H} = - \sum_A^{N_k} \underbrace{\frac{1}{2M_A} \Delta_A}_{\hat{T}_N} - \sum_i^{N_e} \underbrace{\frac{1}{2} \Delta_i}_{\hat{T}_e} + \sum_{A < B}^{N_k} \underbrace{\frac{Z_A Z_B}{r_{AB}}}_{\hat{V}_{NN}} + \sum_{i < j}^{N_e} \underbrace{\frac{1}{r_{ij}}}_{\hat{V}_{ee}} - \sum_A^{N_k} \sum_i^{N_e} \underbrace{\frac{Z_A}{r_{Ai}}}_{\hat{V}_{Ne}} \quad (3.2)$$

where  $Z$  is the nuclear charge,  $M$  is the mass of the nuclei and  $r$  the respective distances between electrons, nuclei or electron and nucleus. The Born-Oppenheimer approximation as introduced in chapter 2 decouples the motion of the nuclei and electrons, due to the larger mass of the nuclei. Thus, the potential energy of the nuclei  $\hat{V}_{NN}$  is treated as a constant.<sup>40, 93, 207</sup>

For a many-electron system, the equation cannot be solved exactly. To find the best approximated energy, the variation principle is applied. It starts with a trial wave function, followed by a variation of this function until a minimum of the energy is obtained:

$$E = \frac{\langle \psi | \hat{H} | \psi \rangle}{\langle \psi | \psi \rangle} \geq E_0 \quad (3.3)$$

#### 3.1.1 Hartree Fock Theory

The electronic wavefunction  $\Psi$  is set up as a product of one-electron wave-functions  $\Phi_i(i)$  ( $i=1, \dots, N$ ) to consider a many-electron system:

$$\Psi(r_1, r_2, \dots, r_N) = \Phi_1(1)\Phi_2(2)\dots\Phi_N(N) \quad (3.4)$$

Taking the Pauli Principle into account, an antisymmetrized product of these one-electron wave functions is used and represented by the Slater determinant:

$$\Psi = \frac{1}{\sqrt{N!}} \begin{vmatrix} \Phi_1(1) & \Phi_2(1) & \dots & \Phi_N(1) \\ \Phi_1(2) & \Phi_2(2) & \dots & \Phi_N(2) \\ \vdots & \vdots & \ddots & \vdots \\ \Phi_1(N) & \Phi_2(N) & \dots & \Phi_N(N) \end{vmatrix} \quad (3.5)$$

This ansatz is used by the *Hartree-Fock* (HF) theory, so that the many-electron wave function is approximated by a single Slater determinant. For the one-electron wave-functions, i.e. the molecular orbitals (MOs) a best set has to be found, that minimizes the expectation value of the Hamiltonian under the constraint of orthonormality ( $\langle \Phi_i | \Phi_j \rangle = \delta_{ij}$ ). Typical, this is done by the Lagrange multipliers. Within this procedure, the HF equations are obtained as follows:

$$\hat{f}(i)|\Phi_i(i)\rangle = \epsilon_i|\Phi_i(i)\rangle \quad (i = 1, \dots, N) \quad (3.6)$$

where  $\hat{f}$  is the Fock operator and  $\epsilon_i$  the orbital energies.<sup>93,207</sup> The Fock operator consists of the one-electron operator  $\hat{h}_i$  and the two-electron operators, the Coulomb operator  $\hat{J}_j$  and the exchange operator  $\hat{K}_j$ .

$$\hat{f}_i = \hat{h}_i + \sum_j^N [\hat{J}_j - \hat{K}_j] \quad (3.7)$$

$$\hat{h}_i = -\frac{1}{2}\nabla_i^2 - \underbrace{\sum_A \frac{Z_A}{r_{Ai}}}_{v^{\text{ext}}} \quad (3.8)$$

$$\hat{J}_j(i)|\Phi_i(i)\rangle = \langle \Phi_j(j) | \frac{1}{r_{ij}} | \Phi_j(j) \rangle |\Phi_i(i)\rangle \quad (3.9)$$

$$\hat{K}_j(i)|\Phi_i(i)\rangle = \langle \Phi_j(j) | \frac{1}{r_{ij}} | \Phi_i(j) \rangle |\Phi_j(i)\rangle \quad (3.10)$$

The one-electron operator  $\hat{h}_i$  describes the motion of one electron in the external potential of all nuclei. The Coulomb operator  $\hat{J}_j$  is defined as the classical electron-electron repulsion (Coulomb repulsion) and the exchange operator  $\hat{K}_j$  takes the effects of spin into account, which originate from quantum mechanics due to the antisymmetry of the Slater determinant.<sup>40,93,207</sup>

The HF equation is solved by the self-consistent field (SCF) procedure, since the HF equation in turns depend on the MOs. In HF, the electron-electron repulsion is treated in an "average" way, i.e. each electron is moving in the electrostatic field of the nuclei and the average field of the other  $N - 1$  electrons. In order to solve the HF equation numerically, the MOs are constructed as a linear combination of atomic orbitals (LCAO):

$$|\Phi_i\rangle = \sum_{\alpha=1}^M c_{i\alpha} |\chi_\alpha\rangle \quad (3.11)$$

where  $c_{i\alpha}$  are expansion coefficients and  $\chi_\alpha$  the atomic spatial orbitals (AO). The set of AOs is called basis set, which contains a Slater type orbital (STO)  $|\chi_\alpha\rangle \propto \exp(-\zeta r)$  or a Gaussian type orbital (GTO)  $|\chi_\alpha\rangle \propto \exp(-\gamma r^2)$ . Minimal basis sets are constructed as STOs, which are in turn linear combinations of simple GTOs. The accuracy of the MOs improves when the number of basis functions increases.

The LCAO ansatz leads to the Roothaan-Hall equation:

$$\mathbf{FC} = \mathbf{SCE} \quad (3.12)$$

where  $\mathbf{F}$  is the Fock matrix,  $\mathbf{S}$  is the overlap matrix,  $\mathbf{C}$  is a  $N \times N$  matrix of LC coefficient ( $c_{i\alpha}$ ) and  $\mathbf{E}$  is a  $N \times N$  matrix of orbital energies ( $\epsilon_i$ ).

$$F_{\alpha\beta} = \langle \chi_\alpha | F | \chi_\beta \rangle \quad (3.13)$$

$$S_{\alpha\beta} = \langle \chi_\alpha | \chi_\beta \rangle \quad (3.14)$$

The Fock matrix ( $\mathbf{F} = \mathbf{H} + \mathbf{G} \cdot \mathbf{D}$ ) contains the one-electron matrix ("core Hamiltonian")  $\mathbf{H}$ , the tensor of the two-electron integrals  $\mathbf{G}$  and the density matrix  $\mathbf{D}$ . To solve the Roothaan-Hall equation, the Fock matrix is diagonalized to get the LC coefficients ( $c_{i\alpha}$ ) and the MO energies.<sup>93,207</sup>

$$F_{\alpha\beta} = H_{\alpha\beta} + \sum_{\gamma\delta} G_{\alpha\beta\gamma\delta} D_{\gamma\delta} \quad (3.15)$$

$$H_{\alpha\beta} = \langle \chi_\alpha | -\frac{1}{2}\nabla^2 | \chi_\beta \rangle - \sum_A Z_A \langle \chi_\alpha | \frac{1}{r_{iA}} | \chi_\beta \rangle \quad (3.16)$$

$$G_{\alpha\beta\gamma\delta} = \sum_{\gamma\delta}^{M_{basis}} \langle \chi_\alpha \chi_\gamma | \frac{1}{r_{ij}} | \chi_\beta \chi_\delta \rangle - \langle \chi_\alpha \chi_\gamma | \frac{1}{r_{ij}} | \chi_\delta \chi_\beta \rangle \quad (3.17)$$

$$D_{\gamma\delta} = \sum_i^{\text{occupied}} c_{\gamma i} c_{\delta i} \quad (3.18)$$

In HF, the tensor  $\mathbf{G}$  is responsible for the time consuming part. So that the computational cost formally scales with  $M^4$ , where  $M$  is the number of basis functions.<sup>40,93,207</sup> To reduce computational time, approximation of  $\mathbf{G}$  are made in semi-empirical methods, see section 3.2.

### Electron Correlation

One drawback of HF is the neglect of electron correlation due to the mean-field approximation, which ignores the direct interaction between electrons. The correlation energy ( $E^{\text{corr}}$ ) is defined as the difference between the exact ground state energy ( $E^{\text{ex}}$ ) and the HF energy ( $E^{\text{HF}}$ ):  $E^{\text{ex}} - E^{\text{HF}} = E^{\text{corr}}$ . The correlation between electrons with opposite spins is called Coulomb correlation, while the correlation between electrons having the same spin is called Fermi correlation. One way to account for the electron correlation is to use a linear combination of Slater determinants:

$$\Psi = c_0 \Psi_0^{\text{HF}} + \sum_{i=1} c_i \Psi_i \quad (3.19)$$

The electron correlations methods differ in how the coefficients  $c_i$  are determined.<sup>40,93,207</sup>

### 3.1.2 Configuration Interaction

*Configuration Interaction* (CI) uses the mentioned ansatz to account for electron correlation by expressing the trial wave function as a linear combination of Slater determinants.

$$\Psi_{\text{CI}} = c_0 \Psi_0^{\text{HF}} + \underbrace{\sum_i^{\text{occ}} \sum_a^{\text{vir}} c_i^a \Psi_i^a}_{\text{S}} + \underbrace{\sum_{i>j}^{\text{occ}} \sum_{a>b}^{\text{vir}} c_{ij}^{ab} \Psi_{ij}^{ab}}_{\text{D}} + \dots \quad (3.20)$$

where  $\Psi_i^a, \Psi_{ij}^{ab}$  are the singly, doubly, etc. excited Slater determinants, which are constructed by exchanging virtual and occupied orbitals. If only single excitation are taken into account the method is called CIS, if additionally double excitations are considered its CISD etc.

Full CI represents all possible excitations.

With the CI method, the ground state wave function, the excited state wave functions as well as the corresponding energies can be determined. As for HF, CI methods are also variational derived to obtain the expansion coefficients and the corresponding energies. The lowest eigenvalue of the CI matrix relates to the CI energy, while the second lowest eigenvalue corresponds to the first excited state etc.<sup>93</sup>

There are some deficiencies and problems of the CI method. Full CI is only useful for very small systems, since the number of determinants grows factorial with the size of the basis set. The methods often used for the calculation of excitation energies are thus CIS and CISD. CIS, however does not improve the ground state energy compared to HF, because of Brillouins Theorem. Additionally, truncated CI methods except for CIS are not size consistent.<sup>40,78,93</sup>

### 3.1.3 Coupled Cluster

*Coupled Cluster* (CC) also takes electron correlation into account, where the wavefunction can be constructed as follows similarly to the CI ansatz, with the exponential operator  $T$ :

$$\Psi = e^T \Psi_{\text{HF}} \quad (3.21)$$

$$e^T = 1 + T + \frac{1}{2}T^2 + \frac{1}{6}T^3 + \dots = \sum_{k=0}^{\infty} \frac{1}{k!} T^k \quad (3.22)$$

The operator  $T^1$  for example accounts for the single excitation, see Eq. 3.23, whereas  $T^2, T^3, \dots$  describe higher excitations.  $T^k$  is then equal to full CI.

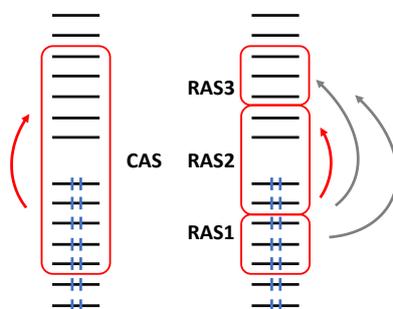
$$T^1 \Psi_0 = \sum_i^{\text{occ}} \sum_a^{\text{vir}} t_i^a \Psi_i^a \quad (3.23)$$

Analogous to CI, CC can also be truncated to reduce computational costs. Truncated CC methods are however size consistent in contrast to truncated CI methods. However, CC is

also a single-reference method and thus can only be used if a single electronic configuration is dominant.<sup>18,40,78,93</sup>

### 3.1.4 Multi-Reference Configuration Interaction

The *multi-reference configuration interaction* (MRCI) method uses a *multi-configuration self-consistent field* (MCSCF) wave function as a reference instead of the single-reference HF wavefunction. In MCSCF not only the coefficients  $c_i$  have to be optimized by the variational principle, but also the MOs used for constructing the determinants. This leads to a complicated variation procedure. To alleviate this computational burden, the energetically lower and virtual higher MOs are frozen and only a part of the MOs is treated actively, the so-called active space. In general the active space is expressed by the *complete active space* (CAS) or the *restricted active space* (RAS), see Fig. 3.2. The RAS is divided into a three-level space: (i) RAS1 includes MOs that are doubly occupied; (ii) RAS2 is equivalent to CAS; (iii) RAS3 includes the non-occupied MOs. In CAS all excitations are possible. The methods are called CASSCF and RASSCF, respectively. An extension to these methods is given by using additional perturbation theory, leading to CASPT2, RASPT2.<sup>40,93,207</sup> CASSCF/CASPT2 is however limited if the active space of the molecule increases. In those cases, for example the simplified multireference method, the so called *Spectroscopy ORiented Configuration Interaction* (SORCI) method was developed for the calculation of optical spectra combining MRCI with perturbation theory, see Ref. 157,187.



**Figure 3.2** Example of complete active space (CAS) and restricted active space (RAS). In CAS( $n,m$ )  $n$  is the number of electrons and  $m$  is the number of active orbitals surrounded with a red box. RAS2 is analogous to CAS, where all excitations are possible. From the RAS1 or RAS2 space only 0, 1 or 2 excitations into RAS3 are possible.

## Density Based Methods

### 3.1.5 Density Functional Theory

In *Density Functional Theory* (DFT), the quantum system is described in terms of electron density. In general, the electron density is the square of the wavefunction. The integral of the electron density over all space gives the total number of electrons  $N$ :

$$N = \int \rho(\mathbf{r}) d\mathbf{r} \quad (3.24)$$

The electron density has maxima (or cusps) corresponding to the position of the nuclei, since the nuclei are point charges. The basic assumption that the electron density determines the properties of a molecule and the energy is given by variation properties are provided by the Hohenberg-Kohn theorems.

The first Hohenberg-Kohn theorem provides a one-to-one mapping between the exact electron density  $\rho(\mathbf{r})$  and the exact external potential  $V_{\text{ext}}(\mathbf{r})$ . The exact external potential  $V_{\text{ext}}(\mathbf{r})$  determines the exact ground state wavefunction and thus, the exact wavefunction is a functional of the electron density  $\psi[\rho]$ . The second Hohenberg-Kohn theorem states that a trial electron density leads to an energy greater or equal to the energy obtained by an exact electron density (Eq. 3.26), which is analogous to the Raleigh-Ritz principle for wave functions.

Thus, the ground state energy can be determined as a functional of the electron density. Hence, the same individual components occur in the ground state energy making also use of the BO approximation:

$$E_0[\rho_0] = T[\rho_0] + E_{\text{ee}}[\rho_0] + E_{\text{Ne}}[\rho_0] \quad (3.25)$$

$$E[\tilde{\rho}] \geq E_0 \quad (3.26)$$

The ground state energy consists of the kinetic energy of the electrons  $T[\rho_0]$ , the energy of the electron-electron interaction  $E_{\text{ee}}[\rho_0]$  and the electron-nuclei attraction  $E_{\text{Ne}}[\rho_0]$ . To make it useful for the description of molecules, the Kohn Sham (KS) approach is used. Kohn and Sham introduce the concept of a non-interacting reference system built from a set of orbitals, i.e. non-interacting electrons in an effective potential  $V_{\text{eff}}$ .<sup>110</sup> Within this approach, a large part of the total energy can be computed exactly. A small part, the non-classical contribution to the electron-electron repulsion will be treated with an approximated functional. Eq. 3.25 can be rewritten as follows:

$$E[\rho] = T_{\text{S}}[\rho] + J[\rho] + E_{\text{XC}}[\rho] + E_{\text{Ne}}[\rho] \quad (3.27)$$

$$E_{\text{XC}}[\rho] \equiv T[\rho] - T_{\text{S}}[\rho] + E_{\text{ee}}[\rho] - J[\rho] \quad (3.28)$$

In Eq. 3.27,  $J[\rho]$  refers to the classical part, the Coulomb interaction between electrons and  $T_{\text{S}}$  to the kinetic energy of the non-interacting reference system, which is thus not equal

to the exact kinetic energy of an interacting system ( $T_s \neq T$ ). The remaining part of the kinetic energy and of the potential energy, the electron exchange are transferred to the so called exchange-correlation functional  $E_{XC}[\rho]$ . Writing Eq. 3.27 more explicitly, the following equation is obtained:

$$E[\rho(\mathbf{r})] = -\frac{1}{2} \sum_{i=1}^N \langle \varphi_i | \nabla^2 | \varphi_i \rangle + \frac{1}{2} \int \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}d\mathbf{r}' + E_{XC}[\rho] + \int V_{Ne}\rho(\mathbf{r})d\mathbf{r} \quad (3.29)$$

The kinetic energy is written in terms of orbitals with the definition, that the density is the square of the Kohn-Sham orbitals,  $\rho(\mathbf{r}) = \sum_{i=1}^N |\varphi_i(\mathbf{r})|^2$ . Also the Coulomb part and the electron-nuclei attraction can be reformulated in an expression containing the orbitals.

At this point, the effective potential  $V_{\text{eff}}$  as well as the orbitals have to be determined. The effective potential  $V_{\text{eff}}$  (Eq. 3.30) contains the potential  $V(\mathbf{r})$  of the nuclei and the Coulomb interaction between one electron and the mean-field created by all other electrons as well as the so-far unknown potential  $V_{XC}(\mathbf{r})$ . The latter corresponds to the exchange-correlation energy  $E_{XC}$  and is defined as the functional derivative of  $E_{XC}$ .

$$V_{\text{eff}}(\mathbf{r}) = V(\mathbf{r}) + \frac{1}{2} \int \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' + V_{XC}(\mathbf{r}) \quad (3.30)$$

$$V_{XC}(\mathbf{r}) = \frac{\delta E_{XC}[\rho]}{\delta \rho(\mathbf{r})} \quad (3.31)$$

Hence, the one-particle eigenvalue equation is given by:

$$\underbrace{\left[ -\frac{1}{2} \nabla^2 + V_{\text{eff}}(\mathbf{r}) \right]}_{\hat{H}_{KS}} \varphi_i = \epsilon_i \varphi_i \quad (3.32)$$

To obtain the orbitals and thus the ground state energy, the Kohn-Sham equations have to be solved iteratively analogous to the HF equations, since the effective potential  $V_{\text{eff}}(\mathbf{r})$  in turn is dependent on the density because of the Coulomb part. As in the Roothaan-Hall equation basis sets are used, where the energy is then obtained by optimizing the coefficients.

As mentioned in section 3.1.1, HF neglects the electron correlation leading to the development of various other methods to take the electron correlation into account. In DFT, the electron correlation is considered by the exchange-correlation functional  $E_{XC}[\rho]$ , which can be calculated using various approximations presented briefly in the following.

The electron correlation in terms of the density is described as a reduced probability of finding an electron in the immediate vicinity of another electron. The correlation of electrons with an opposite spin correlates to the Coulomb hole, while the one of electrons with the same spin to the Fermi hole.<sup>40,93,107</sup>

### Local Density Approximation

The *local density approximation* (LDA) is the simplest method and based on the model of the uniform electron gas or homogeneous electron gas. The  $E_{\text{XC}}[\rho]$  in terms of LDA, is given by:

$$E_{\text{XC}}^{\text{LDA}}[\rho] = \int \rho(\mathbf{r})\varepsilon_{\text{xc}}(\rho(\mathbf{r}))d\mathbf{r} \quad (3.33)$$

$$\varepsilon_{\text{x}}(\rho) = -\frac{3}{4} \left(\frac{3}{\pi}\right)^{1/3} \rho(\mathbf{r})^{1/3} \quad (3.34)$$

where the quantity  $\varepsilon_{\text{xc}}$  is splitted up as a sum of the exchange part and the correlation part, while the latter is not shown here. The exchange part is calculated by the free electron gas model, whereas the correlation part can be determined from quantum Monte Carlo calculations.

LDA reproduces well geometries, however it fails in the case of bond energies due to its overbinding tendency. Hence, LDA does not give a sufficient description of molecules due to the homogeneous electron density.<sup>40,93,107</sup>

### Generalized Gradient Approximation

The *generalized gradient approximation* (GGA) extends the LDA approach by also using the gradient of the density  $\nabla\rho(\mathbf{r})$  to account for non-homogeneity of the electron density.

$$E_{\text{XC}}^{\text{GGA}}[\rho] = \int \rho(\mathbf{r})\varepsilon_{\text{xc}}(\rho(\mathbf{r}), \nabla\rho(\mathbf{r}))d\mathbf{r} \quad (3.35)$$

Several gradient corrected functionals were developed, mostly adding a correction term to the LDA functional. The gradient expansion of the density needs several coefficients, which are not easily determined. Thus, some functionals are created using fitted parameters to reference data. One popular GGA functional using non-empirical parameters is the one by *Perdew–Burke–Ernzerhof* (PBE).<sup>175</sup> There exist also combinations, e.g. the Becke’s (B)<sup>19</sup> GGA exchange functional is combined with the correlation functional of Lee, Yang and Parr (LYP)<sup>122</sup> leading to the functional BLYP.

GGA functionals work better than LDA, however underestimate e.g. reaction barriers. Both LDA and GGA functionals show a self-interaction error.<sup>40,93,107</sup>

### Hybrid Functionals

To cancel the self-interaction error in LDA and GGA, hybrid functionals were developed by adding a certain amount of HF exchange to the GGA exchange-correlation functional. Hence, the exchange-correlation functional can be then computed as:

$$E_{\text{XC}} = \int_0^1 \langle \Psi(\lambda) | V_{\text{xc}}(\lambda) | \Psi(\lambda) \rangle d\lambda \quad (3.36)$$

where  $\lambda$  is defined as the extent of interelectronic interaction, in the range of 0 (none) to 1 (exact). The extension can be generally formulated as  $E_{\text{XC}}^{\text{Hybrid}} = (1 - a)E_{\text{XC}}^{\text{DFT}} + aE_{\text{XC}}^{\text{HF}}$ , where the variable  $a$  differs among the various functionals. One of the most popular hybrid functional is B3LYP given by the following equation:

$$E_{\text{XC}}^{\text{B3LYP}} = (1 - a)E_{\text{x}}^{\text{LSDA}} + aE_{\text{x}}^{\text{HF}} + bE_{\text{x}}^{\text{B88}} + (1 - c)E_{\text{c}}^{\text{LSDA}} + cE_{\text{c}}^{\text{LYP}} \quad (3.37)$$

where the parameters are about  $a \sim 0.2$ ,  $b \sim 0.7$  and  $c \sim 0.8$  obtained by fitting to experimental data. Thus, B3LYP contains about 20% HF exchange. LSDA refers to local spin-density approximation and B88 is a GGA functional.

Hybrid functionals provide accurate geometries and generally perform better compared to LDA or GGA. However, hybrid functionals fail in describing the polarizability of long chains or excitation energies containing Rydberg states or charge-transfer effects in the time-dependent extension.<sup>236</sup>

### Long-Range Corrected Functionals

The asymptotic behavior of the exchange-correlation potential ( $V_{\text{XC}}(\mathbf{r}) = \delta E_{\text{XC}}[\rho]/\delta\rho(\mathbf{r})$ ) has to be described correct, if the described system depends not only on the quality of the KS orbitals but also on virtual orbitals.  $V_{\text{XC}}(\mathbf{r})$  should show a Coulombic decay ( $-1/r$ ) for  $r \rightarrow \infty$ . However, functionals like LDA or GGA fail to describe the right asymptotic behavior of ( $-1/r$ ). Hybrid functionals improve it, however have a too small contribution of the exact exchange. To improve the long-range orbital-orbital interaction, the long-range corrected (LC) functionals divided the Coulomb interaction into a short-range part and a long-range part:

$$\frac{1}{r_{12}} = \underbrace{\frac{1 - \text{erf}(\omega r_{12})}{r_{12}}}_{\text{short-range}} + \underbrace{\frac{\text{erf}(\omega r_{12})}{r_{12}}}_{\text{long-range}} \quad (3.38)$$

where  $r_{12} = |\mathbf{r}_1 - \mathbf{r}_2|$  is the distance of electrons and the error function  $\text{erf}(\omega r_{12})$  is often used as the smoothing function. The parameter  $\omega$  determines then their respective contributions.

LC functionals are thus distinguished through their contribution of the short-range part as described by DFT exchange and the long-range part as described by HF exchange. Thus, the exchange-correlation functional reads:

$$E_{\text{XC}} = E_{\text{Xsr}} + E_{\text{Xlr}} + E_{\text{c}} \quad (3.39)$$

An extension of B3LYP is obtained by the Coulomb-attenuating method (CAM), i.e. CAM-B3LYP, which extends Eq 3.38 with two additional parameters  $\alpha$  and  $\beta$  ( $\text{erf}(\omega r_{12}) \rightarrow [\alpha + \beta \cdot \text{erf}(\omega r_{12})]$ ). This allows a more flexible approach. CAM-B3LYP has therefore an HF exchange of the long-range part of about 60% and 20% HF exchange of the short-range part.<sup>236</sup> The long-range corrected functional to BLYP, LC-BLYP accounts for 100% HF exchange in the long-range part, but has no HF exchange part in the short-range.<sup>209</sup> The extension of

the functional B97 with additionally optimized parameters leads to the functional  $\omega$ B97X, which contains 16% HF exchange at short range and 100% at long-range.<sup>36</sup> Functionals with about full HF exchange in the long-range part show improved idescription of charge-transfer excitations.<sup>46,209</sup>

### 3.1.6 Time-Dependent Density Functional Theory

DFT as described above accounts for properties of the ground state. However, the DFT theory can be extended in a time-dependent way, referred to as TD-DFT, in combination with linear response theory, if excited state properties are of interest.

TD-DFT is conceptionally analogous to KS DFT. In analogy to the first Hohenberg-Kohn theorem, Runge and Gross set up a theorem, which states the existence of a one-to-one mapping between the time-dependent electron density  $\rho(\mathbf{r}, t)$  and the time-dependent external potential  $V_{\text{ext}}(\mathbf{r}, t)$  up to a spatial constant.<sup>186</sup> Hence, since the time-dependent external potential  $V_{\text{ext}}(\mathbf{r}, t)$  determines the time-dependent wavefunction  $\psi(\mathbf{r}, t)$ , the wavefunction is thus a functional of the electron density. To account for the variation principle, the action integral  $A$  is introduced (Eq. 3.40), which is stationary at the exact electron density. The variation of the action integral in terms of the electron density is given by the Euler equation, see Eq. 3.41.

$$A[\rho] = \int_{t_0}^{t_1} \langle \psi(r, t) | i \frac{\partial}{\partial t} - \hat{H}(r, t) | \psi(r, t) \rangle dt \quad (3.40)$$

$$\frac{\partial A[\rho]}{\partial \rho(r, t)} = 0 \quad (3.41)$$

A system with  $N$  non-interacting electrons in an effective potential as introduced above is given in a time-dependent extension as follows by applying  $\rho(\mathbf{r}, t) = \sum_{i=1}^N |\varphi_i(\mathbf{r}, t)|^2$ :

$$\left[ -\frac{1}{2} \nabla^2 + V_{\text{eff}}(\mathbf{r}, t) \right] \varphi_i(\mathbf{r}, t) = i \frac{\partial}{\partial t} \varphi_i(\mathbf{r}, t) \quad (3.42)$$

$$V_{\text{eff}}(\mathbf{r}, t) = V(\mathbf{r}, t) + \frac{1}{2} \int \frac{\rho(\mathbf{r}', t)}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' + \underbrace{\frac{\partial A_{XC}[\rho]}{\partial \rho(\mathbf{r}, t)}}_{V_{XC}(\mathbf{r}, t)} \quad (3.43)$$

The exchange-correlation potential  $V_{XC}(\mathbf{r}, t)$  is thus expressed by the exchange-correlation action  $A_{XC}$ , which similarly collects all unknown remaining parts.<sup>35,46</sup>

### Linear-Response

In order to calculate absorption spectra in a less costly way than full KS TD-DFT, it is extended with linear response scheme developed by Casida.<sup>34</sup> Linear response refers to a perturbation of the first order, i.e. the electric field is treated as a small external perturbation.

Thus, the linear time-dependent response of a time-independent ground state electron density to this time-dependent electric field gives the excitation energy. The density response appears as a peak at a certain frequency related to the excitation energy. Linear response theory introduces a density matrix formalism  $\mathbf{P}$  and an exchange-correlation kernel  $f_{xc}$  to calculate the response in terms of time-dependent perturbation theory. With this ansatz the so called Random-Phase-Approximation (RPA) equations are obtained.

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^* & \mathbf{A}^* \end{pmatrix} \begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix} = \Omega \begin{pmatrix} \mathbf{1} & 0 \\ 0 & -\mathbf{1} \end{pmatrix} \begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix} \quad (3.44)$$

The matrix elements are defined as follows, ( $\omega_{jb\tau} = \epsilon_{b\tau} - \epsilon_{j\tau}$  with  $n_{i\sigma} > n_{a\sigma}$  and  $n_{j\tau} > n_{b\tau}$ , and  $\epsilon =$  molecular orbital energies):

$$A_{ia\sigma,jb\tau} = \frac{\delta_{ij}\delta_{ab}\delta_{\sigma\tau}\omega_{jb\tau}}{n_{j\tau} - n_{b\tau}} + K_{ia\sigma,jb\tau} \quad (3.45)$$

$$B_{ia\sigma,jb\tau} = K_{ia\sigma,jb\tau} \quad (3.46)$$

$$K_{ia\sigma,jb\tau} = \int \Psi_{i\sigma}(\mathbf{r})\Psi_{a\sigma}(\mathbf{r})f_{xc}^{\sigma\tau}(\mathbf{r},\mathbf{r}')\Psi_{j\tau}(\mathbf{r}')\Psi_{b\tau}(\mathbf{r}')d^3\mathbf{r}d^3\mathbf{r}' \quad (3.47)$$

$$f_{xc}(\mathbf{r},\mathbf{r}') = \frac{\delta^2 E_{xc}}{\delta\rho(\mathbf{r})\delta\rho(\mathbf{r}')} \quad (3.48)$$

where the indices (i,j) refer to the occupied orbitals, (a,b) to virtual (unoccupied) orbitals and ( $\sigma,\tau$ ) are the spin indices.  $\mathbf{X},\mathbf{Y}$  are the eigenvectors, which determine the transition density and oscillator strength of a certain excited state.  $\Omega$  represents the associated transition energy.  $K_{ia\sigma,jb\tau}$  is the so called coupling matrix and represents the linear response of the Hamiltonian to the electric field perturbation. The coupling matrix can be formulated for a singlet (S) and triplet (T) case:

$$K_{ia,jb}^S = \int \int \Psi_i(\mathbf{r})\Psi_a(\mathbf{r}) \left( \frac{1}{|\mathbf{r} - \mathbf{r}'|} + f_{xc}[\rho](\mathbf{r},\mathbf{r}') \right) \Psi_j(\mathbf{r}')\Psi_b(\mathbf{r}')d\mathbf{r}d\mathbf{r}' \quad (3.49)$$

$$K_{ia,jb}^T = \int \int \Psi_i(\mathbf{r})\Psi_a(\mathbf{r}) \left( \frac{\delta^2 E_{xc}}{\delta m(\mathbf{r})\delta m(\mathbf{r}')} \right) \Psi_j(\mathbf{r}')\Psi_b(\mathbf{r}')d\mathbf{r}d\mathbf{r}' \quad (3.50)$$

Here, the density is defined as  $\rho = \rho_{\uparrow} + \rho_{\downarrow}$  with the magnetization density  $m = \rho_{\uparrow} - \rho_{\downarrow}$ . The oscillator strength can be then computed with the response matrix eigenvectors and the transition dipole moments.<sup>34,35,61,164</sup>

## 3.2 Semi-Empirical Methods

Semi-empirical methods are approximative methods and can also take electron correlation into account, however in a computational efficient manner. The main approximations are applied to: (i) the core electrons, which are frozen (ii) the valence orbitals, which are described using a minimal basis, i.e. one basis function per valence orbital and (iii) the reduction of the number of two-electron integrals of Eq. 3.15.

The various semi-empirical methods are distinguished by the number of neglected integrals, or parameterized based on high-level quantum calculations or experimental data.

### Wavefunction Based Methods

The main assumption known as the *zero differential overlap* (ZDO) approximation neglects all products of basis functions depending on the same electron coordinate, when located on different atoms. This reduces the overlap matrix  $S$  to a unit matrix:  $S_{\mu\nu} = \delta_{\mu\nu}$ , with  $\delta_{\mu\nu}$  as the Kronecker delta function, which is 0 for  $\mu \neq \nu$  and 1 for  $\mu = \nu$ . The secular equation (Eq. 3.12) is now simplifies to:

$$\mathbf{FC} = \mathbf{CE} \quad (3.51)$$

Furthermore, one-electron integrals are set to zero when three centers are involved, i.e. two from the basis function and one from the operator. And all three- and four-center two-electron integrals are neglected.

The *neglect of diatomic differential overlap* (NDDO) applies the just mentioned approximations. In addition to the approximations done in NDDO, the *intermediate neglect of differential overlap* (INDO) neglects all two-center two-electron integrals, which are not of the Coulomb type, but preserves the one-center integrals. In the *complete neglect of differential overlap* (CNDO) all two-electron integrals are neglected.<sup>40,93</sup>

#### 3.2.1 Zerner's Intermediate Neglect of Differential Overlap

*Zerner's intermediate neglect of differential overlap* (ZINDO) or ZINDO/S (S = spectroscopy) is based on INDO, and was developed for the calculation of electronically excited states, especially for transition metal complexes. The INDO method was chosen instead of CNDO, because one-center exchange is necessary to separate different terms within one configuration. For example, CNDO gives degenerated singlet and triplet states obtained by  $n \rightarrow \pi^*$  transitions. To further increase precision, INDO/S or INDO/1 was used, which takes ionization potentials for parametrization of the one-center core integrals instead of electron affinities to improve spectroscopic results. ZINDO/S includes a CI calculation, i.e. CIS, to calculate the electronic spectrum, also called ZINDO/S-CIS.<sup>15,239</sup>

### 3.2.2 Orthogonalization Model 2

Another semi-empirical method to calculate vertical excitation energies is given by the *Orthogonalization Model 2* (OM2) also referred to as OM2/MRCI. OM2 was developed on the basis of NDDO as introduced above and includes orthogonalization corrections. It is parameterized for the elements H, C, N and O.

Using the Löwdin orthogonalization, the Roothaan-Hall equation (Eq. 3.12) leads to a standard eigenvalue problem:

$$\lambda \mathbf{F} \lambda \mathbf{C} = \lambda \mathbf{C} \mathbf{E} \quad (3.52)$$

where the matrix refers to an orthogonal basis denoted by  $\lambda$ . The overlap matrix  $\mathbf{S}$  becomes the unity matrix. Using orthogonalization corrections on NDDO leads to a reduction of the two-center two-electron integrals, since some of these integrals become very small with an orthogonal basis. The correction based on orthogonalization effects improves the splitting of the MOs, since the ZDO based methods as described above lead to a too low antibonding MO. Thus, the gaps between bonding and antibonding MOs and the resulting excitation energies are underestimated.

The excitation energies are obtained efficiently by applying MRCI using the *graphic unitary group approach* (GUGA), which is an efficient method for evaluating the matrix elements of the CI Hamiltonian. Hence, OM2 makes the selection of a larger active space possible compared to SORCI. Furthermore, all relevant orbitals are included automatically.<sup>113,200,226</sup>

## Density Based Methods

### 3.2.3 Density-Functional Tight-Binding

DFT can treat systems with about 100 atoms, however with increasing number of atoms the computational cost of DFT increases also. The most time consuming step and therefore computational costly is the self-consistent solution of the KS equation (Eq. 3.32), where every matrix element has to be recalculated. Hence, approximations are needed, which are given by the *Density-Functional Tight-Binding* (DFTB) method, the semi-empirical method of DFT.<sup>52,53</sup> It has to be mentioned, that in DFTB, there is no fit to experimental data since the parameterization procedure is based on DFT calculations. The method originates from solid state physics, where the tight binding formalism states that the electrons in a system are approximately tightly bound to the atoms.

The first version of DFTB used a reference density  $\rho_0$  equal to a fixed density leading to a non-self-consistent procedure and simplification of matrix elements which in turns leads to a reduced computational time. This works for some large systems, however, if polarized systems are involved, e.g. molecules with elements of different electronegativities, this assumption fails. Therefore, the reference density  $\rho_0$  is extended with the density fluctuation  $\delta\rho$  around

it, while the reference density  $\rho_0$  is expressed as a sum of neutral atomic densities.

$$\rho(\mathbf{r}) \equiv \rho = \rho_0 + \delta\rho = \sum_a \rho_a^0 + \delta\rho \quad (3.53)$$

The total energy is expanded around the reference density, i.e. the exchange-correlation functional is expanded in a Taylor series:

$$\begin{aligned} E_{\text{XC}}[\rho_0 + \delta\rho] &= E_{\text{XC}}[\rho_0] + \int \left. \frac{\partial E_{\text{XC}}[\rho]}{\partial \rho} \right|_{\rho_0} \delta\rho d\mathbf{r} \\ &+ \frac{1}{2} \iint \left. \frac{\partial^2 E_{\text{XC}}[\rho]}{\partial \rho \partial \rho'} \right|_{\rho_0, \rho_0'} \delta\rho \delta\rho' d\mathbf{r} d\mathbf{r}' \\ &+ \frac{1}{6} \iiint \left. \frac{\partial^3 E_{\text{XC}}[\rho]}{\partial \rho \partial \rho' \partial \rho''} \right|_{\rho_0, \rho_0', \rho_0''} \delta\rho \delta\rho' \delta\rho'' d\mathbf{r} d\mathbf{r}' d\mathbf{r}'' + \dots \end{aligned} \quad (3.54)$$

After some rearrangements of Eq. 3.29, the total energy is given by:

$$\begin{aligned} E^{\text{DFTB}} &= \sum_i^{\text{occ}} \langle \varphi_i | \underbrace{-\frac{\nabla^2}{2} + V_{\text{ext}} + \int \frac{\rho_0'}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' + V_{\text{XC}}[\rho_0]}_{\hat{H}^0} | \varphi_i \rangle \\ &- \frac{1}{2} \iint \frac{\rho_0 \rho_0'}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r} d\mathbf{r}' - \int V_{\text{XC}}[\rho_0] \rho_0 d\mathbf{r} + E_{\text{XC}}[\rho_0] + E_{\text{NN}} \\ &+ \frac{1}{2} \iint \left( \frac{1}{|\mathbf{r} - \mathbf{r}'|} + \left. \frac{\partial^2 E_{\text{XC}}[\rho]}{\partial \rho \partial \rho'} \right|_{\rho_0 \rho_0'} \right) \delta\rho \delta\rho' d\mathbf{r} d\mathbf{r}' \\ &+ \frac{1}{6} \iiint \left. \frac{\partial^3 E_{\text{XC}}[\rho]}{\partial \rho \partial \rho' \partial \rho''} \right|_{\rho_0 \rho_0' \rho_0''} \delta\rho \delta\rho' \delta\rho'' d\mathbf{r} d\mathbf{r}' d\mathbf{r}'' + \dots \\ &= E^0[\rho_0] + E^1[\rho_0, \partial\rho] + E^2[\rho_0, (\partial\rho)^2] + E^3[\rho_0, (\partial\rho)^3] + \dots \end{aligned} \quad (3.55)$$

The different DFTB methods correspond to the truncation of the Taylor series. The truncation after the first-order term results in the original DFTB method. The inclusion of the second-order term leads to DFTB2.<sup>53</sup> And DFTB3 additionally includes the third-order term.<sup>63,64</sup>

In DFTB, the matrix elements are listed in tables so that they do not have to be integrated explicitly for each geometry. Firstly, an LCAO ansatz is used for the KS orbitals, which in turn separates the densities into atomic contributions  $\rho_0(\mathbf{r}) = \sum_i |\varphi_i(\mathbf{r})|^2 = \sum_a \rho_a(\mathbf{r})$ . Secondly, approximations to the matrix elements are done. In general, three-center contributions are neglected. The first term of Eq. 3.55 represents the Hamiltonian matrix elements and is approximated as follows using a minimal basis set of Slater-type for the KS orbitals:

$$E^0 = \sum_i^{\text{occ}} \langle \varphi_i | \hat{H}_0 | \varphi_i \rangle = \sum_i^{\text{occ}} \sum_{\mu} \sum_{\nu} c_i^{\mu} c_i^{\nu} H_{\mu\nu}^0 \quad (3.56)$$

The diagonal elements of the Hamiltonian matrix  $H_{\mu\mu}^0$  leads to the KS eigenvalues, which are calculated by an atomic DFT calculation using the PBE exchange-correlation functional.

The off-diagonal elements  $H_{\mu\nu}^0$  are obtained by a two-center approximation. These matrix elements are listed in tables by calculating different internucleic distances.

The four terms in the second line of Eq. 3.55 depend only on the reference density  $\rho_0$  and are summarized as the repulsive energy  $E_{\text{rep}}$ , which in turn is a sum of two-center potentials.

$$E_{\text{rep}} = \frac{1}{2} \sum_a \sum_b V_{ab}^{\text{rep}}[\rho_a^0, \rho_b^0, r_{ab}] \quad (a \neq b) \quad (3.57)$$

These terms refer to the original DFTB, where no fluctuation of the density was taken into account. The correction terms due to the extension then requires additional approximations to additionally speed up calculations. Thus, the density fluctuation  $\delta\rho$  is expressed as a sum of atomic density fluctuations.

$$\delta\rho = \sum_a \delta\rho_\alpha \quad (3.58)$$

These are described as charge fluctuations  $\delta q_\alpha = q_\alpha - q_\alpha^0$ , where  $q_\alpha$  is the Mulliken charge and  $q_\alpha^0$  is the number of valence electrons of the neutral atom  $\alpha$ . Using this assumption the second order energy term can be rewritten as:

$$E^{2\text{nd}} = \frac{1}{2} \sum_{\alpha\beta} \gamma_{\alpha\beta} \delta q_\alpha \delta q_\beta \quad (3.59)$$

$$\gamma_{\alpha\beta} = \frac{1}{R_{\alpha\beta}} - S(R_{\alpha\beta}, U_\alpha, U_\beta) \cdot h \quad (3.60)$$

where  $\gamma_{\alpha\beta}$  takes the electron-electron interaction into account and is split up into a Coulomb part and an exchange-correlation part.  $\gamma_{\alpha\beta}$  is represented by the integral over a product of two normalized Slater-type spherical charge densities. The resulting values dependent on the nuclear distances  $R_{\alpha\beta}$ , while two limiting cases exist:

- (i) If the distance  $R_{\alpha\beta}$  is large ( $R_{\alpha\beta} \rightarrow \infty$ ), then the function reduces to a Coulomb interaction between the two partial charges.
- (ii) If  $\alpha = \beta$ , then  $\gamma_{\alpha\alpha}$  describes the on-site-self-repulsion, i.e. the electron-electron interaction on atom  $\alpha$ ; therefore  $\gamma_{\alpha\alpha} = U_\alpha$ , where  $U_\alpha$  is the Hubbard parameter, which is the second derivative of the total energy of a single atom with respect to the occupation number of the highest occupied atomic orbital and describes the chemical hardness.

The function  $S$  denotes a short-range function responsible for the correct convergence of  $\gamma_{\alpha\beta}$  at  $R_{\alpha\beta} = 0$ . In Eq. 3.60,  $h$  represents the so called  $\gamma^h$  function and is discussed below in terms of DFTB3. The computation of the charge fluctuation with the Mulliken population analysis in terms of KS orbitals requires the minimization of the energy which leads to the so called *self-consistent charge* DFTB theory (SCC-DFTB).

With these approximations DFTB2 is obtained, which works well comparable to B3LYP and MP2 calculations and performs also well for biological systems.<sup>51</sup> However, in DFTB2 there is a linear relation between the chemical hardness and the atomic size. Further on, the

Hubbard parameter does not depend on the charge fluctuation of the atom, which leads to a poor description of systems, where charged atoms are involved.<sup>63</sup> Hence, in DFTB3 the third order energy term is taken into account to improve also the description of charged molecules. Analogously, approximations are performed leading to the following equations:

$$E^{3\text{rd}} = \frac{1}{3} \sum_{\alpha\beta} \Gamma_{\alpha\beta} \delta q_{\alpha}^2 \delta q_{\beta} \quad (3.61)$$

The desired chemical behavior for charged systems is given by the derivative of the  $\gamma$  function with respect to the charge and expressed by  $\Gamma_{\alpha\beta}$ , (analogous for  $\Gamma_{\beta\alpha}$ , with  $q_{\beta}$  and  $U_{\beta}$ ):

$$\Gamma_{\alpha\beta} = \left. \frac{\partial \gamma_{\alpha\beta}}{\partial q_{\alpha}} \right|_{q_{\alpha}^0} = \left. \frac{\partial \gamma_{\alpha\beta} \partial U_{\alpha}}{\partial U_{\alpha} \partial q_{\alpha}} \right|_{q_{\alpha}^0} \quad (3.62)$$

$$\Gamma_{\alpha\alpha} = \left. \frac{\partial \gamma_{\alpha\alpha}}{\partial q_{\alpha}} \right|_{q_{\alpha}^0} = \frac{1}{2} \left. \frac{\partial \gamma_{\alpha\alpha} \partial U_{\alpha}}{\partial U_{\alpha} \partial q_{\alpha}} \right|_{q_{\alpha}^0} \quad (3.63)$$

The derivative  $\partial \gamma_{\alpha\beta} / \partial q_{\alpha}$  can be calculated analytically and the diagonal term  $(\partial U_{\alpha}) / (\partial q_{\alpha})|_{q_{\alpha}^0}$  is given by the third derivative of the total energy of an atom with respect to the charge.

In organic molecules and especially in biological systems the hydrogen bonds have to be well described. The description of hydrogen bonds is improved by DFTB3 over DFTB2. As mentioned above (Eq. 3.60), the  $\gamma^h$  function was introduced, which is 1 in case of DFTB2. In DFTB3 this function is extended with a parameter  $\zeta$  to improve hydrogen bonding. More details about this function can be found in Ref. 63.

Within these approximations the KS equation are also formulated in matrix form and solved in terms of SCC-DFTB.

$$\mathbf{HC} = \mathbf{SC}\epsilon \quad (3.64)$$

$$H_{\mu\nu} = H_{\mu\nu}^0 + S_{\mu\nu} \sum_c \Delta q_c \left( \frac{1}{2} (\gamma_{\alpha c} + \gamma_{\beta c}) + \frac{1}{3} (\Delta q_{\alpha} \Gamma_{\alpha c} + \Delta q_{\beta} \Gamma_{\beta c}) + \frac{1}{6} \Delta q_c (\Gamma_{c\alpha} + \Gamma_{c\beta}) \right) \quad (3.65)$$

DFTB3 shows for hydrogen bonds an error in binding energy smaller than 3.0 kcal/mol,<sup>63</sup> which is smaller than errors obtained by DFT with standard functionals using medium sized basis sets typically used in QM/MM simulations to e.g. study biological systems. Furthermore, the geometries obtained with DFTB3 improve slightly for charged molecules.<sup>63</sup>

### 3.2.4 Time-Dependent Density Functional Tight Binding

The efficient DFTB method is also extended in a time-dependent (TD-DFTB) approach analogous to the linear response theory of KS DFT to compute excitation energies. The presented method refers to a closed shell system and uses Casida's equation:

$$\sum_{jb} \left[ (\omega_{ia})^2 \delta_{ij} \delta_{ab} + 4 \sqrt{(\omega_{ia}) K_{ia,jb}} \sqrt{(\omega_{jb})} \right] F_{jb} = \Omega^2 F_{ia} \quad (3.66)$$

where  $\omega_{ia} = \epsilon_a - \epsilon_i$  and the indices  $(i, j, \dots)$  refer to the occupied orbitals, while the indices  $(a, b, \dots)$  refer to the virtual orbitals. The coupling matrix  $K_{ia\sigma, jb\tau}$  has to be approximated to

tabulate the two-center integrals, which is done in terms of a Mulliken approximation. In the singlet and triplet case the following matrices are obtained:

$$K_{ia,jb}^S = \sum_{\alpha\beta} q_{\alpha}^{ia} \tilde{\gamma}_{\alpha\beta} q_{\beta}^{jb} \quad (3.67)$$

$$K_{ia,jb}^T = \sum_{\alpha\beta} q_{\alpha}^{ia} m_{\alpha\beta} q_{\beta}^{jb} \quad (3.68)$$

where  $q_{\alpha}^{ia} = \sum_{\mu \in \alpha} q_{\mu}^{ia}$  and  $q_{\mu}^{ia}$  are the Mulliken transition charges.

$$q_{\mu}^{ia} = \frac{1}{2} \sum_{\nu} (c_{\mu i} c_{\nu a} S_{\mu\nu} + c_{\nu i} c_{\mu a} S_{\nu\mu}) \quad (3.69)$$

The two electron integral  $\tilde{\gamma}_{\alpha\beta}$  (see Eq. 3.49) contains then the XC kernel  $f$ , which is evaluated on the full ground state density. However, there are negligible differences between the calculation using the full ground state density or the reference density  $\rho_0$ . Hence,  $\tilde{\gamma}_{\alpha\beta} \approx \gamma_{\alpha\beta}$  and thus the results of SCC-DFTB can be used. The magnetization term  $m_{\alpha\beta}$  corresponds to a short-range nature and is approximated to the quantity  $M$  ( $M = 1/2(\partial\epsilon_{\uparrow}^{\text{HOMO}}/\partial\rho_{\uparrow} - \partial\epsilon_{\downarrow}^{\text{HOMO}}/\partial\rho_{\downarrow})$ ), which in turns is obtained by atomic DFT calculations analogous to the Hubbard parameter. Thus, the excitation energies can be obtained analogous to Eq. 3.44 using these approximations and the corresponding oscillator strength.<sup>164,166</sup>

### 3.2.5 Long-Range Corrected Density Functional Tight Binding

As discussed in section 3.1.5, the long-range corrected functionals improve over standard DFT functionals as the self-interaction error is removed asymptotically through the long range contribution. Hence, the DFTB scheme was also extended to range-separated functionals yielding the LC-DFTB method.<sup>137,165</sup>

The functional used to extend DFTB is based on Baer, Neuhauser and Livshits,<sup>16,129</sup> i.e. splitting the Coulomb interaction ( $\nu$ ) into short-range and long-range:

$$\nu = \frac{1}{r_{12}} = \nu^{\text{sr}} + \nu^{\text{lr}} = \frac{\exp(-\omega r_{12})}{r_{12}} + \frac{1 - \exp(-\omega r_{12})}{r_{12}} \quad (3.70)$$

The functional is dependent on the parameter  $\omega$  and the exchange-correlation functional is given by:  $E_{\text{xc}}^{\omega, \text{DFT}} = E_c^{\text{GGA}} + \alpha E_x^{\omega, \text{LDA}} + E_x^{\omega, \text{HF}}$  (GGA = PBE and  $\alpha = 1$ ).

The formalism of the DFTB method described above has to be changed slightly, since a reference density matrix  $\mathbf{P}_0$  is introduced, as it necessary for the exchange term. Hence, the energy is expanded with a certain reference density matrix and the fluctuation up to the second order. Eq. 3.54 is than rewritten as follows:

$$\begin{aligned} E_{\text{xc}}^{\omega} &= E_{\text{xc}}^{\omega}[\rho_0] + \sum_{\mu\nu} \Delta P_{\mu\nu} v_{\mu\nu}^{\omega, \text{xc}}[\rho_0] \\ &+ \frac{1}{2} \sum_{\mu\nu\alpha\beta} \Delta P_{\mu\nu} \Delta P_{\alpha\beta} f_{\mu\nu\alpha\beta}^{\omega, \text{xc}}[\rho_0] + \mathcal{O}(\delta\rho^3) \end{aligned} \quad (3.71)$$

where  $v^{\omega,xc}$ ,  $f_{\mu\nu\alpha\beta}^{\omega,xc}$  denotes the first and second derivatives of the functional in the AO basis, respectively. The Greek subscripts represent the Slater-type atom-centered basis functions ( $\Phi_\mu$ ).  $P_{\mu\nu}$  is the density matrix, for a closed shell system defined as:

$$\rho(\mathbf{r}, \mathbf{r}') = \sum_{\mu\nu} \underbrace{\left[ 2 \sum_{i=1}^{N/2} n_i c_{\mu,i} c_{\nu,i} \right]}_{P_{\mu\nu}} \Phi_\mu(\mathbf{r}) \Phi_\nu(\mathbf{r}') \quad (3.72)$$

After some rearrangements the total energy can be then written as:

$$\begin{aligned} E = & \sum_{\mu\nu} P_{\mu\nu} H_{\mu\nu}^0 + \frac{1}{2} \sum_{\mu\nu\alpha\beta} \Delta P_{\mu\nu} \Delta P_{\alpha\beta} [(\mu\nu|\alpha\beta) + f_{\mu\nu\alpha\beta}^{\omega,xc}[\rho_0]] \\ & - \frac{1}{4} \Delta P_{\mu\nu} \Delta P_{\alpha\beta} (\mu\alpha|\beta\nu)^{lr} + E_{\text{rep}} \end{aligned} \quad (3.73)$$

The repulsive energy  $E_{\text{rep}}$  depends only on the reference density. For the Hamiltonian  $H_{\mu\nu}^0$  the same two-center approximations are done as described above. The matrix elements are thus also tabulated for different internucleic distances. Using the Mulliken approximation, the four-center integrals are reduced and the two-center integrals become  $(\mu\alpha|\beta\nu)^{lr} = \frac{1}{4} S_{\mu\nu} S_{\alpha\beta} [\gamma_{\mu\alpha}^{lr} + \gamma_{\mu\beta}^{lr} + \gamma_{\nu\alpha}^{lr} + \gamma_{\nu\beta}^{lr}]$ . The integrals  $\gamma_{\alpha\beta}^{lr}$  are then parameterized as in the original DFTB method, however with the long-range contribution. The Hubbard parameter ( $U = \partial^2 E / \partial n^2$ ) in LC-DFTB are then obtained as:

$$U_\alpha^{\text{LC-DFTB}} = \gamma_{\alpha\alpha}^{fr} - \frac{1}{2(2l+1)} \gamma_{\alpha\alpha}^{lr} \quad (3.74)$$

where  $\gamma_{\alpha\alpha}^{fr}$  accounts for the full-range  $\gamma$ -integral and contains the XC-kernel  $f^{\omega,xc}$ . The subscript  $l$  denotes the angular momentum of the highest occupied atomic orbital. Both  $\gamma_{\alpha\alpha}^{fr}$  and  $\gamma_{\alpha\alpha}^{lr}$  contain the range-separation parameter  $\omega$ . With these approximations, the Hamiltonian for the KS equation can be constructed and solved by applying the variation principle.<sup>137</sup>

### 3.2.6 Time-Dependent Long-Range Corrected Density Functional Tight Binding

The time-dependent extension is also applied to LC-DFTB<sup>115</sup> (LC-TD-DFTB) to make the computation of excitation energies possible, since especially range-separated functionals are promising to improve excited state description containing e.g. charge-transfer effects.

As in to section 3.1.5, where linear response theory was described, and similar to the time-dependent extension of DFTB, the total density as well as the magnetization is used. The XC-kernel can than be written as follows:  $f_{\sigma\tau}^{xc} = f^{xc} + \delta_\sigma \delta_\tau \tilde{f}^{xc}$ .

$$f^{xc} = \frac{\partial^2 E}{\partial \rho \partial \rho} \quad \tilde{f}^{xc} = \frac{\partial^2 E_{xc}}{\partial m \partial m} \quad (3.75)$$

Defining  $\mu = \{Alm\}$  and  $\nu = \{Blm\}$ , where  $l$  denotes the angular momentum and  $m$  the magnetic quantum number, the two-center integrals are written as:

$$\gamma_{Al,Bl'}^{\text{fr}} \approx (\mu\mu|v_c + f^{\omega,\text{xc}}|\nu\nu) \quad (3.76)$$

$$\gamma_{Al,Bl'}^{\text{lr}} \approx (\mu\mu|v_c^{\text{lr},\omega}|\nu\nu) \quad (3.77)$$

Using these approximations the coupling matrix can be obtained:

$$K_{ia\sigma,jb\tau} = \sum_{AB} \sum_{ll'} (q_{Al}^{ia\sigma} \Gamma_{Al,Bl'}^{\sigma\tau} q_{Bl'}^{jb\tau} - \delta_{\sigma\tau} q_{Al}^{ij\sigma} \gamma_{Al,Bl'}^{\text{lr}} q_{Bl'}^{ab\tau}) \quad (3.78)$$

where  $\Gamma_{Al,Bl'}^{\sigma\tau} = \gamma_{Al,Bl'}^{\text{fr}} + \delta_{\sigma}\delta_{\tau}\delta_{AB}W_{Al,l'}$ , for further details see Ref.<sup>115</sup> The  $q_{Al}^{ia\sigma}$  are thus the transition charges, as introduced in section 3.2.4. When using local/semilocal functional ( $\omega \rightarrow 0$ ) in Eq. 3.78, the long-range part vanishes and the original TD-DFTB method is obtained. The long-range part introduces an asymmetry of the coupling matrix, requiring the use of a special algorithm by Stratmann *et al.*<sup>205</sup> Thus, the excitation energies are obtained. The oscillator strength can be computed in terms of the Mulliken approximation and the transition dipole matrix. The parameter is set to  $\omega = 0.3/a_0$ , which provides accurate predictions of ionization energies and band gaps from orbital eigenvalues with the ground state LC-DFTB method.<sup>115</sup>

## 3.3 Molecular Dynamics Simulation

### 3.3.1 Molecular Mechanics

The numerical solution of the Schrödinger equation is very time-consuming when large molecules are involved. In case of biological systems, molecular mechanics becomes popular to describe molecules using the laws of classical mechanics. The idea is to describe the energy of the system as a function of the nuclear position, without explicit consideration of the electrons as individual particles. The potential energy is given by empirical force fields.<sup>40,93</sup>

Force fields apply atom specific parameters obtained by experiments or quantum mechanical calculations. The force field energy contains two main parts, which correspond to bonded and non-bonded interactions between atoms. The bonded interactions describe efficiently all of the quantum effects in chemical bonding, the bond, the angles and the torsion of molecules. The non-bonded interactions are characterized as the Van der Waals (VdW) interactions and the electrostatic interaction from Coulomb's law.

The total energy of a force field is summarized as follows:

$$\begin{aligned}
 E = & \underbrace{\frac{1}{2} \sum_i k_i (r - r_i^0)^2}_{E_{\text{bonds}}} + \underbrace{\frac{1}{2} \sum_j k_j^\vartheta (\vartheta_j - \vartheta_j^0)^2}_{E_{\text{angles}}} + \underbrace{\frac{1}{2} \sum_n V_n \cos[n\omega - \gamma_n]}_{E_{\text{torsions}}} \\
 & + \sum_i^N \sum_{j>i}^N \left[ \underbrace{4\epsilon_{ij} \left( \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right)}_{E_{\text{vdW}}} + \underbrace{\frac{1}{4\pi\epsilon_0} \frac{q_i q_j}{r_{ij}}}_{E_{\text{Coulomb}}} \right]
 \end{aligned} \tag{3.79}$$

The potential energy curve of a bond can be described by the Morse potential  $V(r) = D_e[1 - \exp(-\alpha(r - r_0))]^2$ , where  $D_e$  is the depth of the potential energy,  $r_0$  is the equilibrium distance and  $\alpha = \omega\sqrt{\mu/2D_e}$  with  $\mu$  as the reduced mass and  $\omega$  as the frequency of the bond vibration. Force fields use however an harmonic approximation obtained by applying a Taylor expansion around an "equilibrium" bond length  $r_0$ , since for equilibrium geometries the harmonic approximation is satisfied. The Taylor series is truncated after the second order yielding the energy of bonds ( $E_{\text{bonds}}$ ), which represents the first term of Eq. 3.79, where  $k$  is the spring constant.

The second term ( $E_{\text{angles}}$ ) of the total energy in Eq. 3.79 refers to the description of bond angles of molecules. Bond angles can be described analogously by a harmonic potential. Here, the constant  $k^\vartheta$  is related to the bend frequency, while  $\vartheta_0$  is the equilibrium angle.

The dihedral angles are defined by four atoms and describe the rotation around their covalent bonds. The third term of Eq. 3.79 ( $E_{\text{torsions}}$ ) expresses the dihedral angle as a periodic function by a sum of several cosine functions.  $V_n$  is then the amplitude,  $\gamma_n$  the phase shift and  $n$  the periodicity ( $n = 1 : 360^\circ, n = 2 : 180^\circ, n = 3 : 120^\circ$ ).

The last two terms of Eq. 3.79 belong to the non-bonded interactions depending on the distance between the interacting atoms, i.e. the VdW potential and the Coulomb potential. The VdW interactions are described by the Lennard-Jones-(12,6) potential and consists of two parts: (i) the repulsive interaction (Pauli repulsion) of the atoms at short distances described by the  $r^{-12}$  term and (ii) the attractive interaction induced by instantaneous dipole-induced dipole interactions (London interaction) described by the  $r^{-6}$  term. The electrostatic interaction between two molecules is represented as a sum of interactions between pairs of point charges using Coulomb's law. The Coulomb term contains the dielectric constant  $\epsilon_0$ , the distance between two atoms and the respective partial atomic charge  $q$ .

In this work, the *Chemistry at Harvard Macromolecular Mechanics* (CHARMM)<sup>29, 30, 84, 138</sup>

force field developed for proteins and lipids is used. It contains an additional term in Eq. 3.79, the so called Urey-Bradley term ( $\sum_{\text{Urey-Bradley}} = k_u(u - u_0)^2$ ), which describes the angle bending of atom 1 and 3 when three atoms are involved and is a mixture of  $E_{\text{angles}}$  and  $E_{\text{VDW}}$ .  $k_u$  is then the force constant and  $u$  the distance between atom 1 and 3.

Furthermore, several solvent models exist, differing in the bonding description and parameters. In this work, the TIP3P<sup>95</sup> water model is used, where all three atoms are rigid bonded and every atom contains charges, while only the oxygen atom contains Lennard-Jones parameters. Hence, only the oxygen atom can interact.

### Limitations of Force Fields

Force fields can already give a good description of protein structures, however fail when the electronic properties of the system are dominant or need to be accurately described. Hence, processes like the formation or breaking of chemical bonds, charge transfer or photochemical/photophysical processes can not be described accurately by force fields. In those cases, a combination of quantum mechanical methods with molecular mechanical methods (QM/MM) leads to an efficient description, as introduced in the following section.

Another drawback of force fields is the residue specific treatment of amino acids in a protein, i.e. independent of the environment the respective amino acids are treated equally. This can have consequences if polarization effects occur within the environment, which can lead to changes in the charge distribution of the respective amino acid. This is improved when using polarizable force fields, which account for electronic polarization.<sup>238</sup>

Further problems lie in the description of hydrogen bonds, as can be seen when describing the hydrogen bond pattern of the active site of bacteriorhodopsin (bR).<sup>49,225,233</sup> Experimental studies as well as vibrational frequency computations revealed a rigid hydrogen-bond network, as described in chapter 2. Forcefields, however achieve not such a stable hydrogen bond pattern.<sup>14</sup> For further informations, the reader is referred to several e.g. Refs. 47, 123, 211.

### 3.3.2 Quantum Mechanics/Molecular Mechanics

Quantum mechanical methods combined with molecular mechanical methods (QM/MM) leads to an efficient description, which originates by Warshel and Levitt.<sup>223</sup>

In QM/MM only a selected part is treated with QM methods, while the rest of the system is described by MM methods. This QM part is mostly small, for example in the case of photo-activated proteins, only the chromophore and some nearby amino acids or water molecules would be included in the QM region. The total energy of a system within a hybrid QM/MM scheme can be divided into three parts:

$$E = E_{\text{QM}} + E_{\text{MM}} + E_{\text{QM/MM}} \quad (3.80)$$

where  $E_{\text{QM}}$  is the energy of the QM region,  $E_{\text{MM}}$  the energy of the MM region and  $E_{\text{QM/MM}}$  is the interaction of both QM and MM subsystems. The respective QM and MM subsystems

are computed as described in the previous sections, i.e. the MM subsystem using the force field and the QM system by any quantum chemical method, HF, DFT or semi-empirical. The main challenge is to describe the interaction between both subsystems. Several techniques exist to describe this interaction, denoted as embedding schemes.

*Mechanical embedding* describes unpolarized interactions in a simplest way by means of a force field. The bonded as well as the non-bonded interactions between the QM-atoms and the MM-atoms are described as in Eq. 3.79. The QM calculations take place isolated to the QM subsystem. The non-bonded interaction between the QM and MM subsystems is described by VdW interactions. However, there is no Coulomb interaction between the atoms of the QM and MM subsystems. Additionally, the QM charges are not updated during the calculation and the polarization of the wave function by the MM atoms is missing.<sup>31,197</sup>

An improved description of the QM and MM interaction is given by *electronic embedding*, where the MM atoms are allowed to polarize the QM region. Point charges of the MM atoms are included in the Hamiltonian of the QM calculation, and thus the QM atoms feel the electric potential of the MM atoms.

$$\begin{aligned}
 E_{\text{QM/MM}} &= \sum_A^{N_{\text{MM}}} \sum_B^{N_{\text{QM}}} \frac{Q_A \cdot Z_B}{|\vec{R}_A - \vec{R}_B|} - \sum_A^{N_{\text{MM}}} \sum_i^n \frac{Q_A}{|\vec{R}_A - \vec{r}_i|} \\
 &\approx \sum_A^{N_{\text{MM}}} \sum_B^{N_{\text{QM}}} \frac{Q_A \cdot q_B}{|\vec{R}_A - \vec{R}_B|}
 \end{aligned}
 \tag{3.81}$$

The Hamiltonian is computed then by inclusion of the charge  $Q_A$  of the MM atom A and the Mulliken charge  $q_B$  of the QM atom B. The interaction energy  $E_{\text{QM/MM}}$  is then set up by Eq. 3.81 and the VdW part used in the mechanical embedding scheme.

Another improvement is given by *polarizable embedding*, where QM atoms also polarizes the MM region. In this case, a polarizable force field is necessary, which reacts to the changes of the electron density. It requires a double iterative procedure for allowing the electric fields in both the QM and MM regions to be determined in a self-consistent fashion. This method incurs high costs and is therefore only used when high accuracy is needed.<sup>31,197</sup>

The border between the QM region and the MM region in many cases passes a covalent bond. For the study of proteins, amino acids are often included in the QM region, where the QM-MM boundary is then defined between the  $C_\alpha$ - $C_\beta$  bond. Typically, the QM-MM boundary is given by a C-C bond. In such cases, link atoms are introduced to satisfy the valence of the QM region, in most cases the capping atom is a simple hydrogen atom. To avoid overpolarisation the charges of the nearby MM atoms are smeared out.<sup>197</sup>

In this thesis, a DFTB3/CHARMM<sup>51</sup> interface is used, which takes the electronic embedding into account.

### 3.3.3 Geometry Optimization

The protein structure given by a homology model or directly by an X-ray structure has to be geometry optimized for further computations to avoid overstretched bonds or other high energy conformations. For the geometry optimization, several procedures are possible. In general, the gradient of the energy  $\vec{g} = \nabla E(\vec{r})$  ( $\vec{r} = (x_1, y_1, z_1, x_2, \dots, z_N)$ ) is calculated. One important point is the choice of the step size of the gradient descent algorithm. Taking too large steps, one could miss the minimum and a too small step size leads to unnecessarily many calculation steps. One example is the *steepest descent* (SD) method, using the absolute value of the force vector to adjust the step size for the calculation of the energy, i.e. if the gradient is large, the step size is also large.<sup>40,93</sup>

### 3.3.4 Equilibration

The equilibration of the system is provided by different algorithms known as thermostat and barostat. When the temperature is equilibrated an NVT ensemble is generated. The kinetic energy is used to control the temperature, where the initial velocities are obtained by the Maxwell-Boltzmann distribution ( $\langle E_{\text{kin}} \rangle = 3/2NkT$ ). An example is provided by the Nosé-Hoover thermostat, see Ref. 83, 168, which obtains the temperature by an oscillating relaxation. To account for the pressure an NPT ensemble is used. The control of the pressure proceeds analogously to the thermostat and is provided e.g. by the Parrinello-Rahman barostat, see Ref. 173.

### 3.3.5 Molecular Dynamics

The study of biological systems require often sampled structures, which is provided by *molecular dynamics* (MD) simulations. Based on force fields, they are quite efficient to simulate large molecules up to a time scale of microseconds or even longer. To describe the state of a system, the so-called phase space is defined, for N atoms this is a 6N-dimensional space, with three spatial coordinates ( $x_i$ ) and three momentum coordinates ( $p_i$ ). The dynamics of the system is described by a trajectory in the phase space, i.e. space and momentum changes with time ( $\vec{x}_i(t), \vec{p}_i(t)$ ). The initial positions ( $x_i(t_0)$ ) can be obtained from an X-ray structure or a homology model.

As described above, the interaction between the molecules is defined in terms of a potential  $V$  or a force  $F$ , since the force is the negative derivative of the potential with respect to the position  $r$ ,  $F(r) = -\partial V/\partial r$ . The trajectory is then determined by solving Newton's equations of motion ( $F = ma = m\ddot{x}$ ). The time development of the system generates new geometries using a Taylor expansion:

$$x(t_0 + \Delta t) = x_0 + \dot{x}(t_0)\Delta t + \frac{1}{2}\ddot{x}(t_0)\Delta t + \dots \quad (3.82)$$

Several methods exist for numerical integration. One example is the Verlet method, which expands the Taylor series in a "negative" and "positive" time direction. An extension is provided by the velocity Verlet method, which evaluates both  $\vec{x}$  and  $\vec{v}$  in time:

$$\vec{x}(t_0 + \Delta t) = \vec{x}(t_0) + \vec{v}(t_0)\Delta t + \frac{1}{2}\vec{a}(t_0)\Delta t^2 \quad (3.83)$$

$$\vec{v}(t_0 + \Delta t) = \vec{v}(t_0) + \frac{1}{2}\Delta t(\vec{a}(t_0) + \vec{a}(t_0 + \Delta t)) \quad (3.84)$$

where  $\vec{a}(t)$  is the acceleration. Firstly, the positions are calculated, followed by the calculations of the forces and accelerations. Secondly, the velocities are obtained which in turn are used to determine the new positions. Another example of an integrator is "leapfrog", which evaluates the velocities with half-time steps.<sup>40,93</sup>

## 3.4 Spectroscopic Properties

The investigation of biological systems requires several experimental techniques, which in turn require several computational methods to explain experimental results on an atomistic time scale. In this section, firstly an overview of the already presented methods for the computation of excitation energies is given with their generally known advantages and disadvantages. Secondly, the concept of exciton couplings is outlined, as well as their computation with approximative methods. Furthermore, the computation of vibrational frequencies is presented.

### 3.4.1 Excitation Energy

All *ab-initio* and semi-empirical methods introduced in the previous sections have either a time-dependent (e.g. TD-DFT) or variational ansatz (e.g. CIS) for the calculation of excitation energies. Furthermore, the methods are distinguished in single-reference and multi-reference methods. Methods used for the computation of excitation energies in this thesis are:

A) *Single-reference*: (i) linear response: TD-DFT (+LC), TD-DFTB (+LC) and (ii) variational: CIS, ZINDO/S-CIS; and B) *Multi-reference*: OM2/MRCI.

In general, various methods behave often differently when describing excitation energies in vacuum or in the protein environment. *Ab-initio* multi-reference methods (MRCI) yield high accurate excitation energies, however have high computational costs, and it scales unfavorable with the size of the molecule. For the study of biological chromophores MRCI methods are often used as reference method, since these molecules often show a multi-configurational character. OM2/MRCI as a semi-empirical method is therefore advantageous for the study of e.g. retinal proteins, showing a constant shift of 0.3 eV compared to SORCI results.<sup>59,73,82</sup>

CC2 mostly gives accurate absolute excitation energies with an error of about 0.3–0.5 eV, however with high computational cost. It is only applicable, if the transitions are mainly characterized by single reference character. CIS calculations are very fast, however CIS overestimates the excitation energies with large errors of about 0.5–2.0 eV. The main problem

is the lack of correlation energy. ZINDO/S-CIS in contrast performs better than CIS, since it was parameterized to calculate spectroscopic properties.

For TD-DFT methods with reduced computational cost, the quality of the results depends on the functional used for the computation of excitation energies. In general, mainly hybrid functionals can obtain high accurate excitation energies with an error of about 0.1–0.5 eV, however showing systematic errors for certain excitations. It is known, that hybrid functionals fail to describe Rydberg states, valence states of molecules exhibiting extended  $\pi$ -systems or when charge-transfer (CT) states are involved. In these cases long-range corrected (LC) functionals might be promising to improve this failure.<sup>46</sup> TD-DFTB as the semi-empirical variant of DFT can even compute excitation energies three orders of magnitude faster than DFT, however shows the same restrictions as standard DFT, since TD-DFTB is based on the PBE functional.<sup>164</sup> Therefore, LC-TD-DFTB might be also promising, showing already an improved description for a small set of organic molecules including charge-transfer effects with respect to standard DFT functionals as B3LYP.<sup>115</sup> The response to the external field as obtained by the protein environment is also differently treated by each method and is discussed in chapter 4 and 5.

### 3.4.2 Exciton Coupling

Exciton coupling describes the energetic interaction between two molecules which are spatially close, but not in direct physical contact. Spectroscopically, they behave like a single supermolecule. The Hamiltonian of a dimer is constructed as a sum of the monomer Hamiltonians ( $\hat{H}_i, \hat{H}_j$ ) and the electronic interaction  $\hat{V}_{ij}$  between them.

$$\hat{H} = \hat{H}_i + \hat{H}_j + \hat{V}_{ij} \quad (3.85)$$

$V_{ij}$  can be separated into a Coulomb part and an exchange part. The latter depends on the spatial overlap between the two molecular orbitals of monomer  $i$  and monomer  $j$ . An increasing molecular distance leads to an exponential decrease of the wave function overlap. Hence, at a certain distance the exchange interaction can be neglected.<sup>145</sup> The Hamiltonian for a coupled dimer is constructed as:

$$\mathbf{H} = \begin{pmatrix} E_i & V_{ij} \\ V_{ij} & E_j \end{pmatrix} \quad (3.86)$$

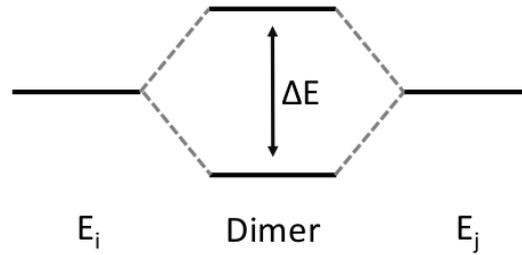
where  $E_i, E_j$  are the excitation energies for each respective monomer. When the exchange interactions are neglected, the off-diagonal elements  $V_{ij}$  are the so-called Coulomb couplings.<sup>41,174</sup> The eigenvalues are obtained by diagonalization of the  $\mathbf{H}$  matrix. If the monomers are oriented parallel and the dimer is symmetric, then  $E_i = E_j$ , and the eigenvalues are:

$$\lambda_{1,2} = \frac{E_i + E_j}{2} \pm V_{ij} \quad (3.87)$$

If the excitation energies of the monomers slightly differ or the monomers are tilted to each other, then  $E_i \neq E_j$  and the diagonalization leads to:

$$\lambda_{1,2} = \frac{E_i + E_j}{2} \pm \frac{1}{2} \sqrt{(E_i - E_j)^2 + 4V_{ij}^2} \quad (3.88)$$

An example, where exciton couplings are computed taking only Coulomb couplings into account, is found for the chromophores in light-harvesting (LH) complexes, where these are separated by about 10-20 Å.<sup>41</sup>



**Figure 3.3** Energy splitting  $\Delta E$  obtained by two coupled molecules.

### Coulomb Couplings

Taking only the Coulomb interaction into account, the off-diagonal element  $V_{ij}$  can be written in terms of transition densities, which is a one-particle density matrix set up by the ground state and excited state wave functions:

$$V_{ij}^{\text{Coul}} = \iint \frac{\rho_i^T(\mathbf{r})\rho_j^T(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}d\mathbf{r}' \quad (3.89)$$

where  $\rho^T$  is the transition density. Several approaches have been developed to calculate the Coulomb coupling, and are used in the study of LH complexes. In most methods, the transition density is not directly used, but approximated as transition dipoles or transition atomic charges. Hence, the coupling results from dipole-dipole interaction or electrostatic interaction through point charges.<sup>43</sup> Some methods used for study of LH complexes are briefly outlined in the following.

The *point-dipole approximation* (PDA) yields the Coulomb couplings through dipole-dipole interactions of two transition dipoles. The PDA calculation requires both the orientation and the magnitude of the transition, which are obtained from absorption spectra of the involved pigments. The orientation of the transition dipole moment is experimentally difficult to determine and often obtained by QM calculations. The Coulomb couplings according to the PDA scheme are computed as:

$$V_{ij}^{\text{PDA}} = |\mu_i||\mu_j| \frac{\kappa}{R_{ij}^3} \quad (3.90)$$

$$\kappa = \hat{\mu}_i \hat{\mu}_j - 3(\hat{\mu}_i \hat{R}_{ij})(\hat{\mu}_j \hat{R}_{ij}) \quad (3.91)$$

where  $|\mu_i|$  and  $|\mu_j|$  represent the dipole lengths and  $R_{ij}$  is the distance between the two dipoles.  $\kappa$  is a geometrical factor and displays the relative orientation between the dipoles.<sup>43</sup> PDA gives a simple representation of Coulomb couplings in terms of dipole strengths, their distance and relative orientation. PDA is only valid for large distances and therefore, it overestimates the Coulomb couplings at short distances and the perpendicular orientation of the two transition dipoles, where  $\kappa$  and thus the couplings become zero.<sup>43, 149, 204</sup>

The *transition charges from electrostatic potential* (TrESP) method computes the Coulomb couplings employing a fit of the atomic partial charges to the electrostatic potential.<sup>139</sup>

$$V_{ij}^{\text{TrESP}} = \sum_{A,B} \frac{q_i(A)q_j(B)}{|\mathbf{R}_A - \mathbf{R}_B|} \quad (3.92)$$

The Coulomb interaction is computed as a sum between all pairs of point charges. For the computation of Coulomb couplings of (bacterio)chlorophyll chromophores, an additional rescaling to the experimental transition dipole moment is done, since the used quantum mechanical methods overestimates the transition dipole moments. QM methods for the calculation of atomic partial charges and transition charges are CIS or TD-DFT/B3LYP using small basis sets.

The *transition density cube* (TDC) works similarly to TrESP and differs in the number of transition charges, since it uses a numerical integration to calculate transition densities represented in 3D grids. The Coulomb couplings are calculated therefore as a sum over all Coulomb interactions of the transition charges in the cube. This represents a large computational effort. As with TrESP, a scaling factor is also used to fit to experimental data.

### DFTB: Computation of Coulomb Couplings

DFTB introduced in section 3.2.3 provides with its time-dependent extension an efficient method to compute excitation energies. The time-dependent extension of LC-DFTB, already showed its improvement to describe organic molecules involving CT excitation energies with respect to standard DFT functionals.<sup>115</sup>

The computation of Coulomb couplings in a spirit of Mulliken transition charges is based on DFTB2 with the time-dependent extension. Taking Eq. 3.66 into account, the transition densities are formulated as follows:

$$N(\mathbf{r}) = \sum_{ia} \sqrt{\frac{2\omega_{ia}}{\Omega}} F_{ia} \phi_i(\mathbf{r}) \phi_a(\mathbf{r}) \quad (3.93)$$

The many-body transition charges are defined as:

$$Q_\alpha = \sum_{ia} \sqrt{\frac{2\omega_{ia}}{\Omega}} F_{ia} q_\alpha^{ia} \quad (3.94)$$

Hence, the Coulomb couplings  $V_{ij}$  are obtained:

$$V_{ij} = \sum_{\alpha \in i, \beta \in j} Q_{\alpha} Q_{\beta} \gamma_{\alpha\beta} \quad (3.95)$$

with the  $\gamma$ -function as defined in section 3.2.3, which describes the second order Coulomb interaction.<sup>114</sup>

The obtained Coulomb couplings underlie the performance of the PBE functional. Following this scheme, J. Kranz has extended it to compute Coulomb couplings based on LC-DFTB, which is used in this thesis for the computation of Coulomb couplings of LH complexes. The reduced computational time makes it also useful for sampling and further extension for the calculation of exciton energy transfer. The advantages of computing Coulomb couplings in the spirit of DFTB is given by the transition charges, which depend then on the accuracy of the DFTB method and can be calculated "on the fly" over a MD trajectory.

### 3.4.3 Vibrational Frequencies

Infrared (IR) or Raman spectroscopy give structural insights into biological systems. The interpretation of IR or Raman spectra is however not always easy for condensed phase so that computational techniques for the computation of vibrational spectra in complex environments are desirable. This is given by the normal mode analysis (NMA) and the Fourier Transform of Time-Correlation Functions (FTTCF).<sup>98</sup>

#### Normal Mode Analysis

The NMA approach employs the harmonic approximation. The diagonalization of the mass-weighted Hessian matrix (Eq. 3.96), which is the second order derivative of the energy with respect to the coordinates, obtains the vibrational modes and frequencies.

$$H = \frac{\partial^2 E}{\partial \Delta \tilde{R}_i \partial \Delta \tilde{R}_j} \quad \text{with} \quad \tilde{R}_i = \sqrt{m_i} \vec{R}_i \quad (3.96)$$

where  $\Delta \tilde{R}$  are the nuclear displacements of atom  $i$  and  $j$  respectively to the equilibrium positions. These are mass ( $m$ ) weighted. The eigenfunctions of the Hessian define the normal modes (vibrational motions) and the eigenvalues correspond to the vibrational frequencies. The computation of the first derivative of the molecular dipole moments in case of IR or the molecular polarizability in case of Raman reveals the spectral intensities. Due to the harmonic approximation the NMA approach describes fast vibrations well but is unsuitable for slow collective motions, where also anharmonic contributions exist. Furthermore, NMA is performed at the equilibrium geometry, which corresponds to 0 K and thus no shape of the vibrational bands can be obtained. To consider also the effect of thermal fluctuation an average over a large number of configurations sampled from equilibrium MD simulations have to be performed.<sup>98, 237</sup>

### Fourier Transform of Time-Correlation Functions

Another method is given by the FTTCF, which is based on Fermi's golden rule and linear response theory leading to the following expressions of the time-correlation functions for the computation of IR from dipole  $\mu$  and Raman spectra from the polarizability  $\alpha$ :

$$\text{IR : } I(\omega) \propto \int_{-\infty}^{\infty} \langle \mu(t) \cdot \mu(0) \rangle e^{-i\omega t} dt \quad (3.97)$$

$$\text{Raman : } I(\omega) \propto \int_{-\infty}^{\infty} \langle \alpha(t) \cdot \alpha(0) \rangle e^{-i\omega t} dt \quad (3.98)$$

From both equations, the frequency, intensity and band shape can be obtained if a time series is available. The extraction of molecular vibration is possible from classical, QM or QM/MM trajectories. Additionally, a correction factor is applied to the intensities to account for the quantum effect on nuclear motion:<sup>22</sup>

$$Q = \frac{\beta \hbar \omega}{1 - \exp(-\beta \hbar \omega)} \quad (3.99)$$

where  $\beta = 1/k_B T$  and  $\omega$  is the vibrational frequency. The combination of FTTCF and SCC-DFTB leads to the possibility to compute vibrational frequencies in the spirit of DFTB, which was already successfully tested for simple chemical systems<sup>237</sup> as well as molecules in complex protein environments.<sup>178</sup> The formalism of the combination of both FTTCF and DFTB can be found e.g. in Ref. 98. A benchmark study by Welke et al.<sup>228</sup> using DFTB3 with the 3ob parameter set for the vibrational analysis of carboxylic acids revealed a mean absolute error of  $6 \text{ cm}^{-1}$  with respect to PBE for the frequencies of the  $\nu_{COOH}$  band, and  $7 \text{ cm}^{-1}$  for the shifts due to hydrogen bonding interaction with  $\text{CH}_3\text{OH}$ . Hence, the 3ob parameters are also used in this thesis for the FTIR computation.

---

## Benchmark of LC-DFTB on Retinal and Bacteriochlorophyll

### 4.1 Introduction

The chromophores in rhodopsins and light-harvesting (LH) complexes exhibit characteristic absorption spectra. The maxima and shape of these absorption spectra result from different influences, which are responsible for shifting, shoulders or broadening of the absorption spectra compared to the single chromophore in vacuum. The shifting of these absorption spectra is also defined as "color-tuning". The influencing factors for color-tuning are:

(i) *chromophore structure*

Retinal, the chromophore of rhodopsins and bacteriochlorophyll (BChl), the chromophore in LH complexes can show differences when comparing the structures within the protein environment with those in vacuum. Steric interactions in the protein matrix could lead to a twist of retinal,<sup>3</sup> while BChl can lose the planarity of the ring system leading to a bowl shape.<sup>212</sup> This has a certain indirect influence on the absorption energy due to the change of the ground state geometry.

(ii) *interaction with the electrostatic environment, direct hydrogen-bonding or Van der Waals (VdW) interactions*

The electrostatic interactions of the chromophore with the environment due to ionic, polar or polarizable groups of amino acids can "tune" the absorption energy. Additionally, a specific hydrogen bond network involving the chromophore influences also the absorption spectrum, since it may lead to charge transfer and VdW interactions.<sup>128,235</sup>

(iii) *exciton coupling*

LH complexes contain several BChl chromophores, which interact with each other. Excitonic

couplings have therefore also a significant impact on the absorption spectrum, since they depend on the arrangement and the distances of the chromophores.<sup>147</sup>

The explanation of the absorption spectra is experimentally not always fully available, which makes computational methods necessary to have a deeper insight on an atomistic scale. Computational methods have to deal with all these influences and additionally with the size of the chromophore as well as the respective system when investigating the absorption spectra in biological systems. They can disentangle the factors, responsible for the shifts of the absorption maxima and the shoulders and band width of the absorption spectra. However, the choice of the method is important in order to confirm that the explanation of the absorption spectra results from correct physical description and not from error cancellation.

### 4.1.1 Computational Investigation of Color-Tuning

Previous studies on retinal have shown,<sup>222</sup> that not all methods are apt to estimate all the influencing effects in the right magnitude making them useless for the computational study of color tuning in biological systems. The reasons are the respective errors, which may appear in different ways. Even though, methods can compute absolute excitation energies with high accuracy, there is no better accuracy than 0.1 eV compared to experimental data, which is nevertheless a huge error. The error in absolute energies has to be systematic, so that it can be scaled in order to be comparable to experimental values.

The first criterion for the study of color-tuning is, that computational methods have to give accurate ground state geometries, since they correlate with the excitation energy. Geometrical parameters like the bond-length alternation (BLA), i.e., the difference of neighboring single and double bond lengths of conjugated molecules, represents one measure factor to see the structural differences and the influence on the excitation energies. The BLA correlates with the C=C stretch frequency and thus with the absorption maximum.<sup>82,108</sup> The BLA values however vary significantly with the computational method. Therefore, the precise value of the BLA has a sizable impact on the excitation energy. In some cases the ground state geometry cannot be determined with the same method as used for the calculation of the excitation energies. One method could give accurate ground state geometries, however could lead then to large errors in the computation of excitation energies, or vice versa.

The second criterion is, that computational methods should also accurately describe the response of the excitation energy to external electrostatic fields. Hence, the inclusion of the protein environment is also important for the study of color-tuning. Errors occur, when computational methods can give an accurate description of excitation energies in vacuum, but fail when including the electrostatic environment.

The third criterion are influences by interaction of several chromophores as found in LH complexes. The resulted exciton couplings are also a challenge for computational methods

in a biological environment, since not only the chromophore interaction itself but also the influence of the environment has to be taken into account.

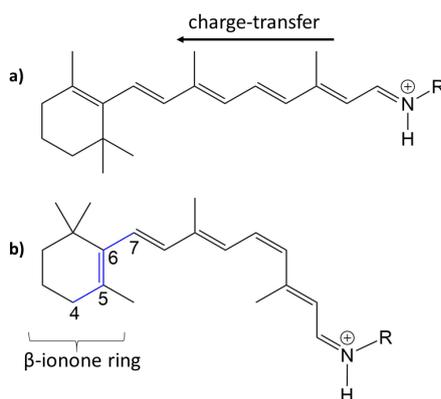
Furthermore, especially in biological systems using *ab-initio* and TD-DFT approaches, excitation energies are often computed for single structures, e.g. for QM/MM optimized geometries of chromophore and protein environment. However, often sampling is required for an improved description of chromophore and protein environment fluctuations. Therefore, semi-empirical methods are necessary to be able to compute several of thousands structures. A deeper introduction about the importance of sampling is given in chapter 5. In the following, a brief overview of previous computational studies for both retinal and BChl is given.

#### 4.1.2 Excitation Energy of Retinal

Retinal has a long polyene chain with a  $\beta$ -ionone ring and a Schiff base (SB), see Fig. 4.1. In rhodopsins, retinal is covalently bound via a SB to a lysin of the apo-protein. Therefore, the retinal is positively charged forming a protonated SB (RSBH<sup>+</sup>). The excitation of retinal leads to a change in bond order and to a shift of the positive charge along the polyene chain, i.e. the excited state of retinal has a strong charge-transfer character.

In general the spectral characteristics of retinal absorption are mostly affected by the protonated state of the SB, the planarity of the conjugated  $\pi$ -system of the retinal, the distance of the protonated SB to the negatively charged residues, and by the electronic environment of the retinal.<sup>59, 82, 220</sup>

In organic solvents the absorption maximum of the chromophore retinal is about 450 nm,<sup>130</sup> while in proteins it varies from 360 to 635 nm.<sup>108</sup> The protein environment can drastically modulate the absorption maximum of the chromophore due to the high electronic polarizability and structural flexibility of retinal.



**Figure 4.1** Retinal bonded via a Schiff base to the apo-protein (R=LYS), here R=H. a) all-*trans* configuration and b) 11-*cis* configuration, dihedral angle of the  $\beta$ -ionone ring highlighted in blue, which varies dependent on the twist.

## Previous Computational Studies

The electronic states of retinal shows a multi-reference character requiring the use of multi-reference configuration interaction (MRCI) methods. CASSCF/CASPT2 describes retinal excited state properties in agreement with experimental results, however with high computational costs.<sup>12,56,191</sup> CASPT2 energies for single conformations or potential energy surface (PES) evaluated on CASSCF geometries, however, lead to poor results, since CASSCF overestimates the BLA of retinal.<sup>8,218,220</sup> Here, geometry optimization e.g. at the DFT level of theory leads to an improvement. An alternative method for the study of retinal is the *ab-initio* spectroscopy oriented configuration interaction (SORCI) method, which yields results with the same accuracy than CASPT2 results, however with reduced computational costs.<sup>187,222</sup> Due to the lack of gradients, DFT geometries are required as well.

High accurate single-reference methods such as CC2 can describe retinal excited states for a mostly planar retinal configuration or without the inclusion of the  $\beta$ -ionone ring.<sup>193-196</sup> The optimization of the retinal geometry in the ground and excited state is only accurate if far away from the conical intersection region.<sup>195,214</sup> Hence, CC2 is not a reliable method to study e.g. the isomerization of retinal.

HF as optimization method overestimates the BLA significantly leading thus to higher absolute excitation energies in retinal.<sup>222</sup> The computational low-costly single reference method CIS leads to an overestimation of the excitation energies and reacts also sensitive to the multi-configurational character.<sup>220</sup>

DFT methods can be used to describe ground state properties of retinal. While GGA-functionals underestimate the BLA, hybrid functionals predict the BLA better with an increasing amount of exact exchange included.<sup>220</sup> For excited states, TD-DFT with GGA functionals fail due to the charge transfer character, variation of excitation energy with geometry and external electrostatic field are not well predicted, in some cases even qualitatively wrong.<sup>220</sup>

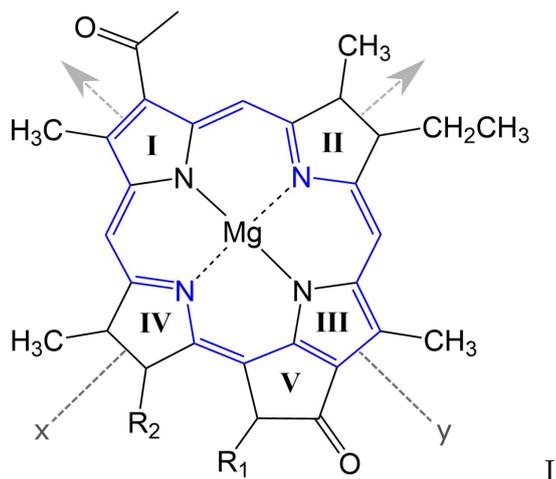
The weaknesses of standard DFT functionals might be improved by the long-range corrected (LC) functionals, as previous benchmark studies on organic molecules involving charge-transfer excitations have shown.<sup>89,90</sup> A previous study concerning the isomerization of retinal using CAM-B3LYP as LC functional showed an improved description of the ground and first excited state PES compared to standard DFT functionals.<sup>185</sup> Another study investigating the excited state relaxation of retinal showed that LC-BLYP provides results in accordance to CASPT2.<sup>215</sup>

Methods with low computing costs based on semi-empirical Hamiltonians are used for sampling excitation energies. In our previous work on retinal protein, we have used OM2/MRCI, since its performance for relative excitation energies is comparable to SORCI,<sup>222,227</sup> while absolute excitation energies are blue-shifted.

### 4.1.3 Excitation Energy and Exciton Coupling of BChl a

Bacteriochlorophyll a (BChl a) is one of the naturally occurring BChl's as introduced in chapter 2. It possesses bacteriochlorin as skeleton, which in turn is a derivative of porphyrin but with reduced  $\pi$ -electrons. Thus, the aromaticity is not distributed over the whole ring system. The absorption spectrum of BChl a is characterized by the  $Q_x$ ,  $Q_y$  and the Soret band. In this study, only the  $Q_y$  excitation energy is considered. The transition dipole moments are lying on the plane of the ring system, where the one of the  $Q_y$  lies along the axis of pyrrol ring I to III,<sup>69,70</sup> see Fig. 4.2.

BChl a occurs in both LH complexes, the FMO and LH2 complex, which are used for the benchmark and introduced in chapter 2. In organic solvents, BChl a shows an absorption maximum of about 770 nm (1.61 eV)<sup>213</sup> for the  $Q_y$  excitation energy. In proteins the BChl a chromophores exhibit different absorption maxima dependent on the LH complex, which in generally are red-shifted. For example, in the FMO complex the absorption spectrum is found in a range of about 790 to 825 nm (1.50-1.57 eV).<sup>210</sup> In the case of the LH2 complex two absorption maxima are found of about 800 nm (1.55 eV) and 850 nm (1.46 eV).<sup>240</sup>



**Figure 4.2** Bacteriochlorophyll a (BChl a); in blue diaza[18]-annulene substructure; full model:  $R_1 = \text{COOMe}$ ,  $R_2 = \text{phytyl-tail}$ ; truncated model:  $R_1 = \text{H}$ ,  $R_2 = \text{CH}_3$ ; In gray the positions of the transition dipole moments of the  $Q_x$  and  $Q_y$  excitation energies.

### Previous Computational Studies

Several studies have investigated the excited states of BChl chromophores until today. Due to the large amount of different methods used, for example the  $Q_y$  excitation energy vary in a range of 300-1000 nm (4.13-1.24 eV) dependent on the computational method.<sup>125</sup> To get an overview about the state of the art in the literature, the computational methods used until now are summarized in the following.

### a) Excitation Energies

A benchmark study on BChl a chromophores of the FMO complex investigating the behavior of several methods (CIS, ZINDO/S-CIS, CC2 and TD-DFT) on BChl a excited state properties, found a multi-configurational character of the BChl a ground state. Therefore, they used DFT/MRCI as reference method.<sup>127</sup> RASSCF/RASPT2 was also used for the investigation of BChl a chromophores occurring in the LH2 complex,<sup>9,10</sup> choosing an active space of 25  $\pi$ -electrons. However, the study of BChl a in terms of MRCI is time consuming and requires a lot of computational effort due to large active space. Furthermore, the MRCI methods cannot be performed as "black box" methods, so that the choice of the taken orbitals have to be done carefully. MRCI methods can be used as reference, but are impractical for the study of LH complexes due to their high amount of chromophores. Both mentioned studies revealed small shifts of about 0.003-0.045 eV for the LH2 complex<sup>9</sup> and 0.020-0.075 eV for the FMO complex<sup>127</sup> when including environmental effects.

High accurate single reference methods such as CC2 was found to react sensitive on the multi-configurational character of the BChl a chromophores compared to DFT/MRCI.<sup>127</sup> A study on BChl structures using CC2 and the algebraic diagrammatic construction through second order, denoted as ADC(2) however showed, that both methods are applicable when applying on a truncated BChl model.<sup>206</sup> Hence, CC2 or ADC(2) can be used as reference method for a reduced model, which is practical, since it does not require the manually choice of active orbitals. However, it is not usable for the study of BChl chromophores as occurring in proteins, due to the sensitivity on their multi-configurational nature.

The computational cheaper single-reference method CIS is also not applicable for the study of BChl excited state properties, since it tends to overestimate the excitation energies. Additionally, it reacts also sensitive on the multi-configurational character of the BChl a chromophores.<sup>125,127</sup>

Density based methods became popular in the study of BChl chromophores, because of the reduced computational costs and being robust to the multi-configurational character of the BChl a chromophores when using hybrid or long-range corrected (LC) functionals.<sup>125,127,159</sup> GGA functionals give similar trends to hybrid functionals or DFT/MRCI when considering the isolated chromophore but is affected by state mixing. Including the environmental effects, GGA tends to underestimate the excitation energies.<sup>127</sup> Problems with hybrid functionals occur, if one wants to investigate also the carotenoides, which are present for example in the LH2 complex. Here errors occur describing the excited state properties of the long polyene chain, where LC functionals might improve better.<sup>160</sup> The LC functional CAM-B3LYP was applying for a benchmark study on the BChl a chromophores in the FMO complex, as well as on an investigation of solvation effects. Both studies concluded, that CAM-B3LYP performs similar or better than hybrid functionals. In the case of the FMO complex, CAM-B3LYP could reproduce the same trend as the reference method DFT/MRCI.<sup>127</sup> In the study with

different solvents, CAM-B3LYP gives an improved description of the electronic structures of the ground and excited states compared to B3LYP with respect to the investigation of electrostatic potentials.<sup>81</sup>

In a semi-empirical manner, mainly ZINDO/S-CIS is preferred for the study of LH complexes.<sup>39,125,171,172</sup> However, the benchmark study on the FMO complex mentioned above figured out, that ZINDO/S-CIS tends to overestimate environmental effects.<sup>127</sup>

## b) Exciton Coupling

LH complexes contain several BChl a chromophores, which could be responsible for shifting or broadening of the absorption spectra due to exciton coupling. Since the chromophores are spatially not very close together, i.e. distances are about 10-20 Å the excitonic couplings can be approximatively described by Coulomb couplings.

As introduced in chapter 3, several approximative methods were developed for the study of Coulomb couplings. Mostly all employ scaling techniques to fit the experimental data. The studies, summarized here are directly performed on the crystal structures after adding hydrogen bonds of the respective proteins, i.e. there should be less structural deviations.

Methods used for the computation of excitonic couplings in LH complexes are e.g. PDA, TrESP, TDC or on the basis of subsystem TD-DFT, the so called frozen-density embedding (FDE) approach.<sup>43,91,111,160</sup> In literature, the reason for color-tuning and broadening of the absorption spectrum of the FMO complex is referred more to chromophore-protein interactions than to chromophore-chromophore interactions.<sup>41,147</sup> Studies investigating exciton couplings in LH complexes were thus focusing mostly on the LH2 complex, where the B850 ring was set to be reasonable for color tuning due to the closely arranged chromophores.

In the appendix in Tab. A.3 and Tab. A.4 several studies on the LH2 complex are listed, differing in the used method for the computation of excitation energies and excitonic couplings. Thus, the overall comparison of the studies is not easy. However, they can be compared separately focusing e.g. only on the exciton couplings. The studies are performed on the crystal structure of the LH2 complex. The strongest couplings vary between 0.02–0.10 eV depending on the method.

In previous studies, the PDA method was set to overestimate couplings when compared to TrESP or TDC, especially for distances smaller than 15 Å.<sup>101,149,204</sup> Regarding Tab. A.3 in the appendix, the Coulomb coupling values using TrESP, TDC or FDEc are lying in a range between 0.03–0.04 eV, while those using supermolecular calculations based on CIS and INDO/CIS are in a range of 0.07–0.10 eV. TrESP and TDC are using a scaling factor to fit the experimental data of the transition dipole moment. The study using INDO/S-CIS however pointed out, that the experimentally determined transition dipole moments vary of about 6.1–7.7 D, dependent on the solvent.<sup>39</sup> Hence, one has to pay attention, that methods scaled on the experimental value depends thus on the accuracy of the experiment. Unscaled

TDC calculations obtained also higher values of about 0.07 eV.<sup>192</sup> The study on CIS based supermolecule calculations concluded additional, that short-range effects, i.e. exchange contribution are also involved, however relatively weak by an amount of half of the magnitude of the B800-B800 couplings, which are in turns about one-tenth of the B850-B850 couplings.<sup>190</sup>

The summary of the previous computational studies showed, that standard TD-DFT functionals fail when describing excitation energies where charge-transfer effects are involved. Even though, previous studies on both retinal and BChl proteins showed, that LC functionals lead to an improved description over standard functionals,<sup>42,127,185,192,204,215</sup> there is still no systematic benchmark regarding several influencing factors of color tuning. Semi-empirical methods are necessary to compute sampled chromophores to account also for the shape of the absorption spectrum. The semi-empirical TD-DFTB method, however faces the same problems as TD-DFT methods using GGA functionals, since TD-DFTB is based on the PBE exchange-correlation expression.<sup>164</sup>

Recently, a LC functional was implemented into the density functional tight binding (LC-DFTB) method and was extended to the time-dependent scheme (LC-TD-DFTB). It is about three orders faster than the LC-TD-DFT method and therefore a promising method for the study of absorption spectra of rhodopsins and LH complexes.<sup>115,165</sup> Especially for LH complexes, where the computation of the exciton couplings is a computational bottleneck, an accurate and fast LC-TD-DFTB method would be of great help in computational studies. LC-TD-DFTB was already successfully benchmarked for a test set of small organic molecules, where charge-transfer excitations were involved.<sup>115</sup>

Therefore, in this chapter a broad and systematic benchmark is performed to determine the accuracy of LC-TD-DFT and LC-TD-DFTB on retinal and BChl chromophores for the description of several influencing factors on their absorption maxima. Firstly, the benchmark study of LC-TD-DFT and LC-TD-DFTB on excitation energies is presented with respect to geometry effects and electrostatic effects. Secondly, a benchmark of LC-TD-DFTB for the computation of exciton couplings in the case of BChl is performed. The performance of LC-TD-DFT and LC-TD-DFTB in rhodopsins and LH complexes is presented in chapter 5.

## 4.2 Computational Details

### 4.2.1 Models and Geometry Optimization

The three retinal model systems used in section 4.3 are displayed in Fig. 4.1 and 4.6 and were taken from Ref. 222. For the study of the chromophore bacteriochlorophyll a (BChl a), geometry optimizations were carried out using the Turbomole program package.<sup>2</sup> For the geometry optimization, HF and DFT, employing the functionals BH-LYP, B3LYP and BLYP together with the Karlsruhe basis sets def2-SV(P), def2-SVP, and def2-TZVP were used. Additionally, geometry optimizations were performed at the SCC-DFTB level of theory,<sup>53</sup> denoted DFTB in the following, and at the DFT level of theory with the range-separated CAM-B3LYP functional together with the def2-SVP and def2-TZVP basis, as implemented in the ORCA program package.<sup>158</sup>

To account for protein environment effects on the excitation energies of BChl a, the Fenna-Matthews-Olson (FMO) complex (PDB code 3EOJ) of *Prosthecochloris aesturii* was taken from Ref. 127.

For the benchmark study of the exciton couplings of the BChl a chromophores, an artificial system was set up. The artificial system contains a truncated BChl a structure (see Fig. 4.12), which was first optimized by SCC-DFTB and then used to build a dimer. Several dimers were built differing in distances and rotation. Furthermore, the exciton coupling benchmark was extended by using a ring system of BChl a as occurring in biological systems. The B850 ring system of the light-harvesting complex II (LH2) (PDB code 1LGH) from *Rhodospirillum rubrum* was taken and QM/MM optimized. The QM/MM optimizations were performed using the DFTB3/3OB method for the QM region as implemented in the GROMACS package (version 2017.1).<sup>1,118</sup>

The QM region of the LH2 complex contains one BChl a chromophore, respectively. The phytyl tail was not included in the QM region, since the QM/MM bond is defined between the CP1 and CP2 atoms. CP1 is the first C atom after the double bond of the phytyl tail, see Fig. 2.6 in chapter 2. The neglect of the phytyl tail in the QM region saves computational cost and does not significantly effect the excitation energies, cf. Ref. 39,125. The system was QM/MM minimized using the steepest descent (SD) algorithm (100000 steps with a tolerance of  $1000 \text{ kJ mol}^{-1} \text{ nm}^{-1}$ ).

### 4.2.2 Excitation Energies

The optimized geometries were subsequently used for the calculation of vertical excitation energies at HF/CIS, ZINDOS/CIS, TD-DFT (BP86, B3LYP) and LC-TD-DFT (CAM-B3LYP,  $\omega$ B97X, LC-BLYP) as implemented in the ORCA program package.<sup>158</sup> ZINDO/S is short hand notation for ZINDO/S-CIS and LC-TD-DFT is abbreviated as LC-DFT in the following. The vertical excitation energies resulted by using geometries obtained with the def2-SVP basis

were used in the present work, as exemplary comparison to results obtained from geometries optimized using the def2-TZVP basis showed no significant difference.

For the HF/CIS computations, the def2-SVP basis set was chosen. For the ZINDO/S calculations an active space of (10,10), i.e., including the 10 highest occupied and the 10 lowest unoccupied states, was employed, which yields a sufficient agreement with experimental results for both BChl a and retinal.<sup>131,172</sup>

The Tamm-Dancoff approximation (TDA) has been invoked for all TD-DFT calculations and the Resolution of Identity (RI) to speed up the computations. The def2-TZVP basis set was used with RIJCOSX for the Coulomb integral and HF exchange and def2/J as an auxiliary basis set in the RI approximation. In order to avoid large errors due to possible charge-transfer contributions in vertical excitations, long-range corrected functionals such as CAM-B3LYP,  $\omega$ B97X and LC-BLYP were used. These functionals are compared in the present work as they differ in the amount of exact exchange. For example, LC-BLYP contains 0% exact exchange at short range and 100% at long-range,<sup>209</sup>  $\omega$ B97X contains 16% exact exchange at short range and 100% at long-range<sup>36</sup> and CAM-B3LYP consists of 19% exact exchange at short range and 65% at long-range.<sup>236</sup> Vertical excitation energies were furthermore performed using the time-dependent generalization of DFTB (TD-DFTB)<sup>166</sup> without and with long-range corrected functionals (LC-TD-DFTB).<sup>115,137,165</sup> LC-TD-DFTB is abbreviated as LC-DFTB in the following.

In order to provide reference values for vertical excitation energies, wavefunction methods are often applied. Since in the present study large systems are investigated, only low-scaling wavefunction methods could be applied to guarantee applicability for the entire set of models used. Therefore, the reference method for small molecules is a second-order approximate coupled-cluster singles and doubles (CC2) method and the algebraic diagrammatic construction scheme of second order, denoted ADC(2).<sup>76,77</sup> Both methods employ the resolution of the identity (RI). Since with increasing system size even these efficient excited-state methods can become unfeasible, these two methods are also employed in the scaled-opposite-spin approximation in combination with the Laplace transformation (LT) to reduce the computational scaling, i.e. LT-SOS-RICC2 and LT-SOS-RIADC(2),<sup>229,230</sup> In the following, however, these methods are denoted SOS-CC2 and SOS-ADC(2) since the resolution-of-the-identity and the Laplace transformation introduce deviations only significantly below the method error. All these wavefunction based methods are performed by S. Höfener using the Turbomole program package.<sup>2</sup>

### 4.2.3 Exciton Coupling

Supermolecule calculations were performed on the artificial system using CIS and TD-DFT with the functionals B3LYP,  $\omega$ B97X and CAM-B3LYP as implemented in the ORCA program package.<sup>158</sup> For the CIS calculations the def2-SVP basis set was applied, while for the DFT

functionals the def2-TZVP basis set was applied. Additionally, supermolecule calculations were performed using TD-DFTB and LC-DFTB. Coulomb couplings were calculated for the artificial system using LC-DFTB and TrESP. The computation with the TrESP method were performed by S. Maity using no scaling factor. The partial transition charges of the TrESP method are based on B3LYP and CAM-B3LYP, with the def2-TZVP basis set.

For all BChl a dimers of the B850 ring supermolecule calculations were performed using the LC-DFT functional CAM-B3LYP as well as LC-DFTB. Analogous to the artificial system, the Coulomb couplings were computed using LC-DFTB and TrESP, while the TrESP calculations were performed by S. Maity. For the biological system, the standard frozen TrEsp procedure was applied, i.e. one individual pigment in gas phase was taken and the equilibrium position was calculated to subsequent determine the TrEsp charges. The partial transition charges are based on CAM-B3LYP. This is a common method for the application on trajectories, and therefore also applied here for the biological system. The BChl a chromophores were QM/MM optimized as described above.

## Results and Discussion

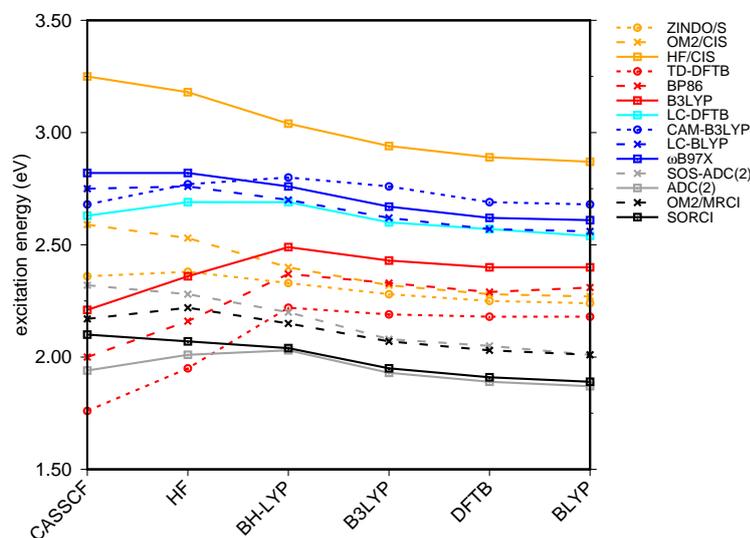
### 4.3 Retinal

#### 4.3.1 Bond Length Alternation

Variations of the retinal conformations can have significant influences on the absorption shift, since the ground state geometry correlates with the excitation energy. As described in section 4.1.1, one geometrical parameter is given by the BLA. The correlation to the excitation energy is reflected in a hypsochromic (blue) shift when the BLA is increased.

The BLA differs significantly between the computational methods used to optimize the geometry of retinal. While BLYP has the smallest BLA, it increases in hybrid approaches with the amount of exact exchange added. HF and CASSCF lead to the highest BLA.

The use of different methods for geometry optimization allows therefore to assess the impact of the BLA on the excitation energy in a systematic way. Additionally, several methods are used for the computation of the excitation energies. In Fig. 4.3 the excitation energies vs. BLA of the optimized all-*trans* retinal geometries<sup>222</sup> are plotted. The excitation energies and the BLA values are given in the appendix, see Tab. A.1.



**Figure 4.3** Influence of the optimization method upon excitation energies in eV; Results from BP86, DFTB, B3LYP, OM2/CIS, HF/CIS, OM2/MRCI, and SORCI are taken from Ref. 222. Results from SOS-ADC(2) and ADC(2) are taken from Ref. 86.

SORCI as the reference method displays an expected blue shift with an increasing BLA of the retinal structures.<sup>222</sup> Similarly, CIS-based methods show a blue shift of the excitation

energies with an increased BLA, however, HF/CIS largely overestimates the excitation energies by a factor of about 1.5. SOS-ADC(2) follows the same trend as SORCI, whose absolute excitation energies are in the range of OM2/MRCI. ADC(2) shows also the same trend as SORCI, however obtains a red shift for CASSCF and HF optimized structures (-0.16 eV). A similar trend but with a significant red shift for HF and CASSCF optimised retinal structures (-0.37 eV) is obtained by the deficient behavior of GGA and hybrid TD-DFT methods as discussed in Ref. 222. This may be related to a deficiency of TD-DFT with GGA and hybrid functionals describing excited states structures. Starting an geometry optimization near the Franck Condon (FC) point should lead to a geometry, where the ground state BLA is inverted, i.e., double bonds become single bonds and vice versa. This behavior is correctly described by CASSCF/CASPT2 methods, but neither by TD-DFT<sup>222</sup> nor by CC2.<sup>193</sup> Here, the bond alternation is increased, instead of inverted.

The *ab-initio* LC functionals, LC-BLYP and  $\omega$ B97X, where the HF exchange part of both functionals belonging to the long-range part is 100% show the same trend as SORCI, while CAM-B3LYP with only 80% HF exchange still has a qualitative problem. The same trend as described by CAM-B3LYP is also obtained by LC-DFTB, however showing a significant improvement for the HF and CASSCF optimized retinal structures compared to DFT GGA and TD-DFTB based on GGA functional. As expected, absolute excitation energies are overestimated by LC functionals.

An important property is the change in excitation energy due to variation of the BLA, which shows up e.g. in the width of the absorption band due to fluctuations. SORCI has a range of  $\Delta = 0.21$  eV. SOS-ADC(2) overestimates this range, while ZINDO/S as well as CAM-B3LYP and LC-DFTB underestimate it. A similar range compared to SORCI is obtained by LC-BLYP as well as  $\omega$ B97X.

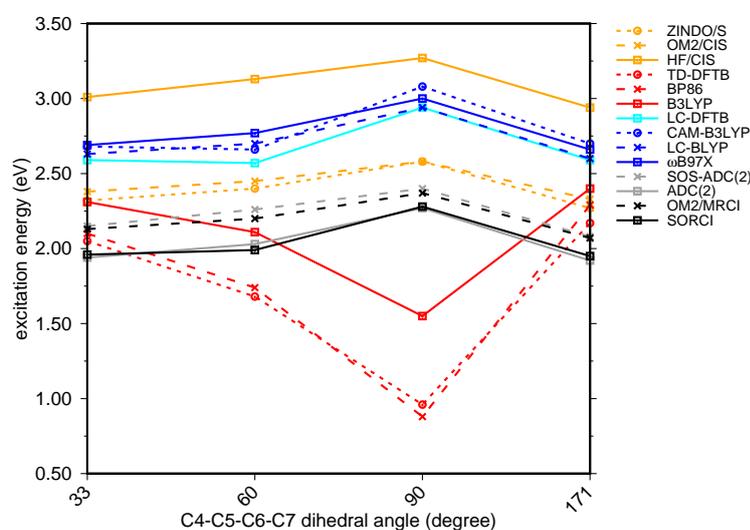
### 4.3.2 Twist of the $\beta$ -Ionone Ring

Another important geometrical parameter of retinal is the dihedral twist angle of the  $\beta$ -ionone ring, see Fig. 4.1. A planar retinal conformation describes a  $\pi$ -system being delocalized over the whole conjugated chain, whereas a highly twisted C<sub>6</sub>-C<sub>7</sub> single bond leads to a  $\sigma - \pi$  mixing in the natural orbitals. This leads to a blue shift of the excitation energy with respect to the planar structure.

To investigate the effect of delocalization on excitation energies, a 6-*s-cis*-11-*cis* PSB model with different dihedral twist angles is taken from Ref. 222. The results of the methods, which are tested, are shown in Fig. 4.4 and are listed in Tab. A.2 in the appendix. Despite their failure for the BLA, ADC(2) and SOS-ADC(2) properly reproduce the reference method SORCI, showing a blue shift of the excitation energies for the twisted retinal conformation. CIS based methods follow also the same trend as SORCI with higher absolute excitation energies as obtained in the case of the optimized all-*trans* retinal structures in Fig. 4.3. The

failure of standard DFT functionals is visible in a red shift of the excitation energies by the retinal twist. As for the BLA, LC-DFT functionals as well as LC-DFTB show an improvement obtaining a blue shift for the twisted retinal configuration.

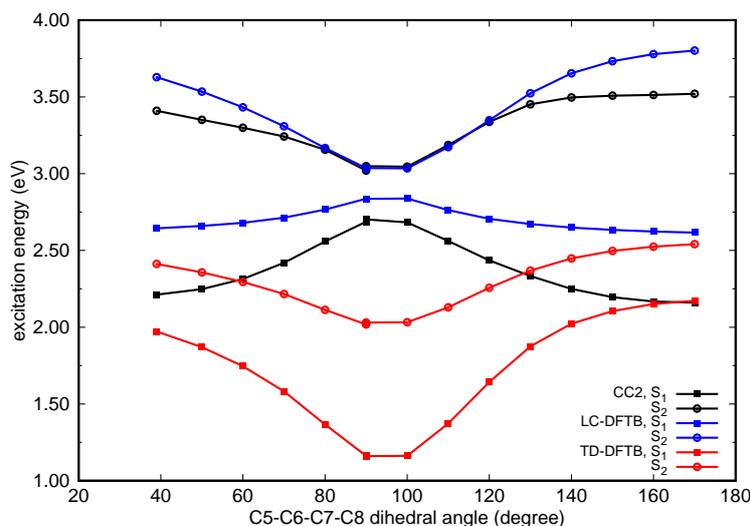
This failure of hybrid and GGA TD-DFT approaches is even more dramatic as in the case of BLA. In many retinal proteins, i.e. in bovine rhodopsins a blue shift of the absorption energy is found due to a twist in the retinal conformation around the C<sub>6</sub>-C<sub>7</sub> single bond.<sup>3</sup> This seems to be one efficient way of color tuning, induced by steric interaction of the environment with the retinal chromophore. It is therefore encouraging to see, that the LC-DFT methods are able to reproduce this feature.



**Figure 4.4** Excitation energies in eV of a 6-*s-cis*-11-*cis* PSB in vacuum for different twist angle. Results from BP86, DFTB, B3LYP, OM2/CIS, HF/CIS, OM2/MRCI, and SORCI are taken from Ref. 222. Results from SOS-ADC(2) and ADC(2) are taken from Ref. 86.

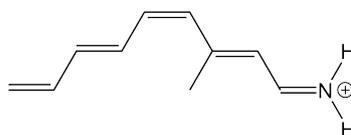
### Potential Energy Surface (PES)

Additionally, LC-DFTB is benchmarked for the description of the potential energy surface (PES) from the 6-*s-cis*-11-*cis* PSB to a 6-*s-trans*-11-*cis* PSB configuration. The ground-state pathway is taken, optimized with PBE0 by M. Wanko, and the energy profiles of the first two excited states are calculated using LC-DFTB. Fig. 4.5 compares LC-DFTB with CC2 results from Ref.,<sup>48</sup> where the same optimized geometries were used. LC-DFTB describes qualitatively the correct PES as CC2, similar to the CAM-B3LYP functional not shown here, which corrects the wrong behavior of B3LYP displaying a red shift of the potential energy surface of the first excited state, cf. Ref. 48. This failure of standard TD-DFT functionals is shown with TD-DFTB, see Fig. 4.5.



**Figure 4.5** Potential energy curves ( $S_1$  and  $S_2$ ) for the ground-state isomerization coordinate from 6-*s-cis*- to 6-*s-trans*-PSB11. The CC2 values are taken from Ref. 48.

### 4.3.3 Effect of the Electrostatic Environment



**Figure 4.6** PSB5.

Retinal proteins absorb light over a wide range of nearly 300 nm, and a major contribution results from electrostatic interactions with the protein environment. Negatively charged groups, i.e. counterions in the binding pocket of the retinal chromophore close to the protonated Schiff base ( $\text{RSBH}^+$ ) have a prominent contribution of the electrostatic environment causing a significant hypsochromic shift of the  $S_1$  excitation energy.<sup>56</sup> Dependent where the counterions are located, e.g. near the  $\text{RSBH}^+$  or near the  $\beta$ -ionone ring, the energy gap varies. The location of the counterion near the  $\text{RSBH}^+$  energetically stabilizes the electronic ground state and leads thus to a larger energy gap, which in turn results in a blue shift.<sup>11</sup> The location of the counterions near the  $\beta$ -ionone ring leads then to a red shift due to the energetical stabilization of the excited state.

For meaningful benchmarks, a simple model has been proposed focusing on the  $S_1$  vertical excitation energy of a PSB5 model, see Fig. 4.6 interacting with a point charge in its immediate environment. The PSB5 geometry was reoptimized with DFTB using a charge of -1.1 au

at the distance of 2.22 Å to the SB proton,<sup>222</sup> which results in a shift when computing the excitation energy of the reoptimized model compared to the model in vacuum. The reason is the point charge near the protonated SB, which lower the energy level of the ground state. Additionally, the resulted shifts are computed by including the point charges and reported in Tab. 4.1.

**Table 4.1** Excitation energy shifts in eV of the PSB5 model due to the (i) reoptimization with the point charge (refers to the column 0.0) and (ii) the computation including the point charge. In all cases, the excitation energies were computed at the geometry optimized in the presence of a -1.1 point charge. Results from ADC(2) and SOS-ADC(2) are taken from Ref. 86.

Charge <sup>a</sup>	VEE <sup>b</sup>	Shifts		
		0.0	-0.5	-1.1
ZINDO/S	2.78	+0.02	+0.13	+0.31
CAM-B3LYP	3.34	-0.01	+0.10	+0.27
LC-BLYP	3.30	+0.03	+0.18	+0.37
$\omega$ B97X	3.36	+0.04	+0.18	+0.37
LC-DFTB	3.24	+0.03	+0.15	+0.28
ADC(2)	2.61	+0.04	+0.28	+0.60
SOS-ADC(2)	2.74	+0.09	+0.35	+0.71
SORCI <sup>c</sup>	2.67	+0.03	+0.28	+0.58
OM2/MRCI <sup>c</sup>	2.61	+0.07	+0.39	+0.73
TD-DFTB <sup>c</sup>	2.77	-0.09	+0.00	+0.11

<sup>a</sup>Vertical excitation energy.

<sup>b</sup>Charges in atomic units.

<sup>c</sup>Taken from Ref. 222.

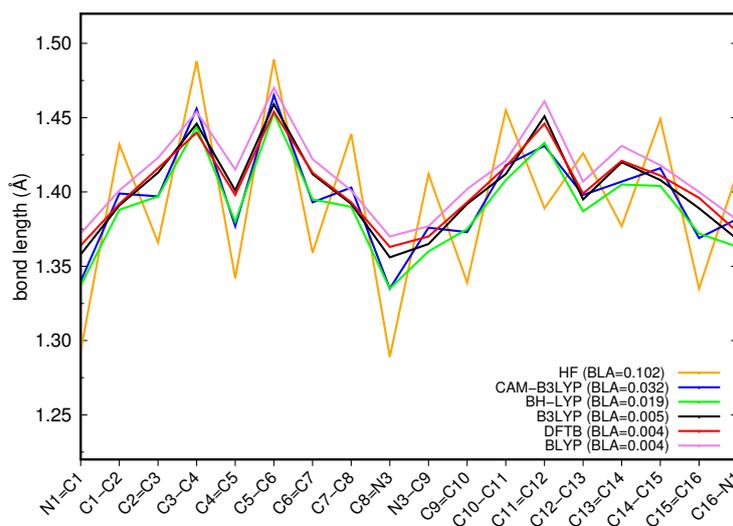
SORCI, considered to be the reference method, shows a blue shift of +0.58 eV with the charge -1.1 au. ADC(2) and SOS-ADC(2) nicely reproduce this shift, while TD-DFT using GGA and hybrid functionals grossly underestimate the shift finding values of less than 0.2 eV.<sup>222</sup> TD-DFTB reproduces this failure, and ZINDO/S yields values only half of the SORCI values. An improvement is observed with LC-DFT functionals as well as LC-DFTB. LC-BLYP and  $\omega$ B97X show with +0.4 eV the largest shift in contrast to CAM-B3LYP and LC-DFTB (+0.3 eV). However, the shifts are still substantially smaller, which implies a weakness of LC functionals. This can be denoted short-hand as "color-weakness". Therefore, TD-DFT with GGA and hybrid functionals are not able to predict color-tuning effects, and also LC-DFT can not be regarded as a quantitative method in this respect.

## 4.4 Bacteriochlorophyll

### 4.4.1 Bond Length Alternation

Studies concerning the excitation energies of BChl a are mostly performed on the crystal structure or on QM/MM optimized structures. The protein environment, however has already an influence on the structure of the chromophore, which can have drastic influences on the computed excitation energies due to the different ground state geometries. To overcome such problems in this benchmark study, one BChl a structure was optimized in vacuum using different methods, i.e. HF, DFT (GGA, hybrid and LC functionals) and DFTB.

The BLA of the optimized BChl a structure is calculated to figure out the structural differences. For the BLA, the diaza[18]-annulene substructure is chosen visualized as blue bonds in Fig. 4.8, as previously used for studies of porphyrin.<sup>6</sup> The measured bond lengths and the numbering can be found in the appendix, see Fig. A.1.

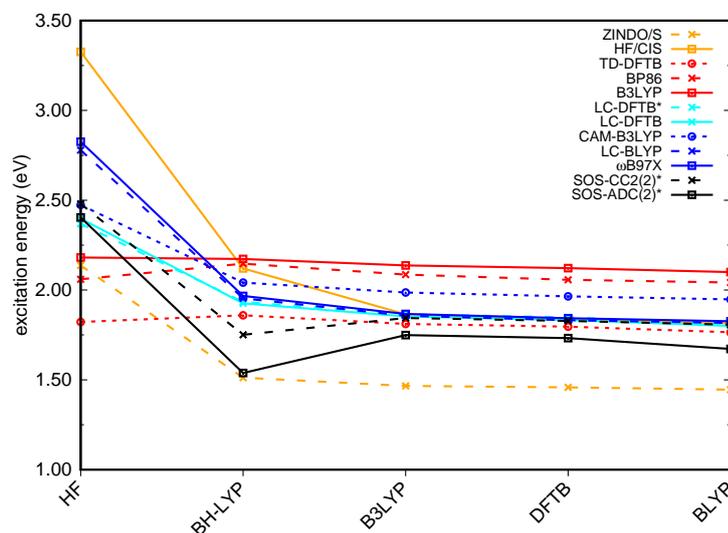


**Figure 4.7** Bond lengths of BChl a in vacuum obtained by different optimization methods and BLA values in Å.

In Fig. 4.7, the bond lengths are plotted as well as the BLA values (in Å). HF significantly shows in contrast to all other methods another trend of the bond lengths, i.e. an alternation of the single and double bonds over the whole ring. Standard DFT functionals (B3LYP and BLYP) as well as DFTB receive less alternations reflected also by the small BLA ( $\sim 0.004$ ) compared to HF (0.102). A previous study reported that B3LYP leads to an improved description of BChl a geometries compared to HF, which is shown also here by a significantly overestimation of the BLA by HF.<sup>125</sup> DFTB reveals a similar BLA than B3LYP and is therefore

applicable for e.g. QM/MM MD simulations of LH complexes. The more HF amount the functional possesses the stronger is the alternation of bond length as seen in case of BH-LYP (0.019) or CAM-B3LYP (0.032) showing a higher BLA value than B3LYP. CAM-B3LYP as LC functional displays also deviations due to the higher BLA and the additional obtained bowl shape instead of a planar ring structure of BChl a compared to all other methods. Interestingly, this bowl shape can be also found in proteins, which will be focused on in section 4.4.2 and in chapter 5.

These ground state geometries are now used for the computation of excitation energies to see the influences of the different structures, see Fig. 4.8. The values are listed in Tab. A.5 in the appendix. The use of a multi-reference method like SORCI as it is the case for retinal is here not feasible, because of the large active space of BChl a required for an accurate description. Hence, the wavefunction-based methods SOS-CC2 and SOS-ADC(2) are employed to provide excitation energies which do not suffer from the self-interaction error. However, these methods are sensitive to the multi-reference character of the ground-state wavefunction. The  $D_1$  value is on one hand a measure for the single-reference character and thus the applicability of these methods for a given case<sup>92</sup> but on the other hand this  $D_1$  value is not a strict measure. Following Ref. 206, we have truncated some substituents, see Fig. 4.2, which breaks the symmetry of the BChl a chromophore, in order to reduce the  $D_1$  values. When comparing the excitation energies of the truncated and the full geometries using LC-DFTB, we find that the truncation does not lead to significant changes, see Tab. A.5 in the appendix.



**Figure 4.8** Excitation energy (eV) of BChl a in vacuum represented as a function of the excited state method; \*truncated structure. Results from SOS-CC2(2) and SOS-ADC(2) are taken from Ref. 86.

In Fig. 4.8, all methods show for standard DFT as well as DFTB optimized methods few variations in the excitation energies, which is in accordance with the small variation of the BLA. SOS-CC2 and SOS-ADC(2) exhibit a significant blue shift in case of the HF geometry (+0.8 eV) and a small red shift for the BH-LYP optimized structure. CIS yields a blue shift in the case of those geometries which exhibit an increased BLA. A similar trend is observed in the case of ZINDO/S, but the absolute excitation energies are slightly reduced. The GGA and hybrid TD-DFT functionals underestimate the excitation energies in the case of the HF geometry, similar to the behavior in the case of retinal. For those geometries with a small BLA, GGA and hybrid functionals provide an accurate description for BChl a for both the ground-state geometry and subsequent excitation energy calculations.<sup>159</sup> For the geometries which exhibit an increased BLA, however, these functionals are not sufficient, while LC-DFT or LC-DFTB obtain a (significant) blue shift analogous to the reference methods SOS-CC2 and SOS-ADC(2). Comparing the LC functionals and LC-DFTB among each other,  $\omega$ B97X and LC-BLYP reveal a similar trend showing slightly larger excitation energy values for the HF optimized structures since both functionals have a higher HF amount. LC-DFTB exhibits a similar trend to CAM-B3LYP with lower absolute excitation energies.

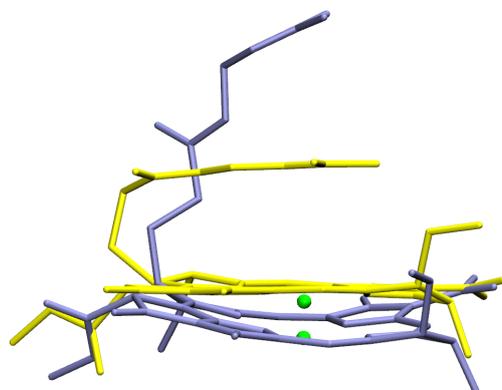
The CAM-B3LYP optimized structure is also used for the computation of excitation energies with the above mentioned methods, however not plotted in Fig. 4.8 since a bowl shape is obtained and not a planar BChl a ring shape as received by the other methods. The CAM-B3LYP structure leads in the case of SOS-CC2 and SOS-ADC(2) a red shift of the excitation energies, which is not obtained by the other methods and might be due to the shape. CIS and TD-DFT functionals with a higher amount of HF, i.e.  $\omega$ B97X and LC-BLYP obtain a blue shift. This observation is further discussed in the next section, when including the protein environment.

#### 4.4.2 Effect of the Electrostatic Environment

The protein environment manifests itself in the change of the BChl a structure due to steric interactions with the surrounded protein and in a shift of the excitation energies due to the electrostatic field. The magnitude of the shift however in LH complexes is expected to be small as shown in section 4.1.3.

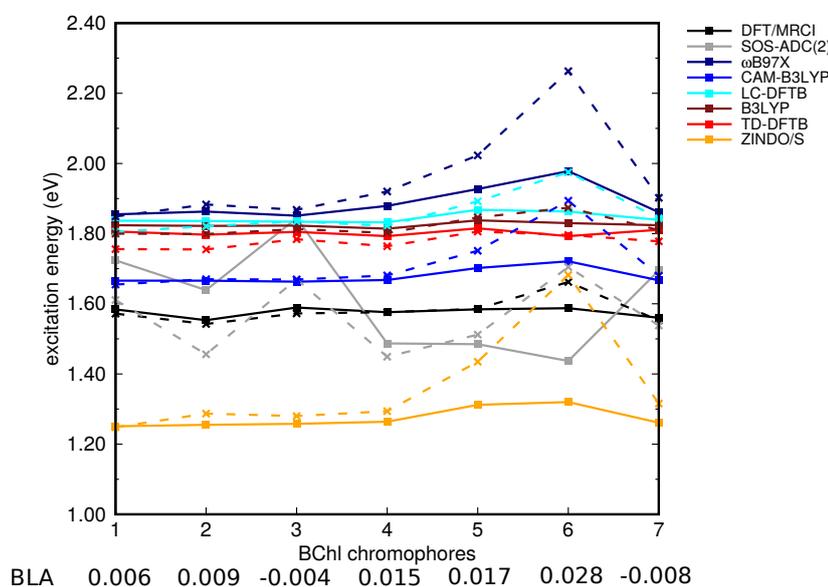
In general, BChl a displays in proteins not always a planar shape, but also a bowl shape as also obtained with the CAM-B3LYP optimized structure in vacuum, as shown in Fig. 4.9. This was suggested to be caused by the out-of-plane displacement of the Mg ion due to the axial ligand coordinated to the central Mg ion "pulling" it out of the plane.<sup>212</sup> The influences of the axial ligand or other hydrogen bonded amino acids to the BChl a chromophore seems however to have few impact on the excitation energy, cf. e.g. Ref. 9, 127, 159. A deeper discussion on the LH2 complex is given in chapter 5. In contrast to the LH2 complex, the shift range of the absorption spectrum of the FMO complex compared to the single BChl a chromophore in

organic solvent was set to be dominated by protein-chromophore interactions.<sup>147,160</sup> Hence, in this section, the FMO complex is used for the benchmark of the protein environmental influences on the excitation energies described by LC-DFT and LC-DFTB.

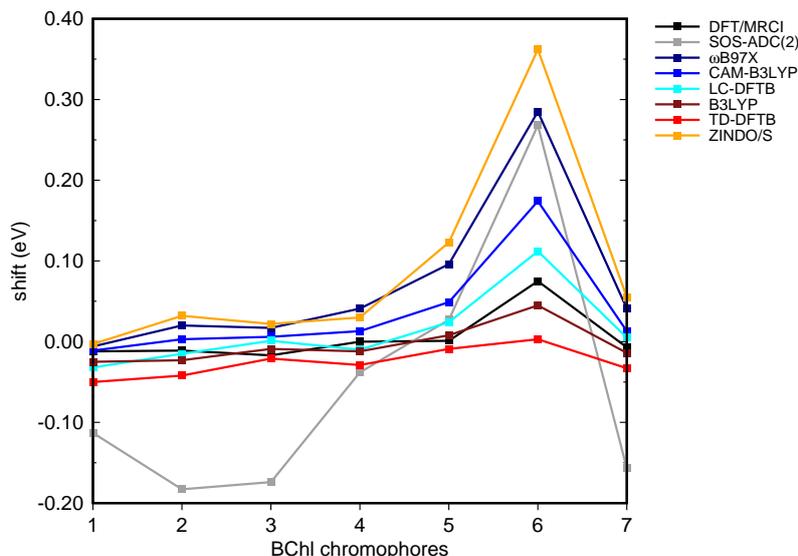


**Figure 4.9** Two BChla of the FMO complex, yellow: planar ring shape; blue: bowl ring shape.

The QM/MM optimized structures from the FMO complex of *P.aestuvarii* from Ref. 127 are thus taken to assess the accuracy of the influence of the electrostatic field upon excitation energies. Since CC2 fails to provide accurate excitation energies for the full BChla geometry, we use the DFT/MRCI values from Ref. 127 as reference values in the following.



**Figure 4.10** Excitation energies of the FMO complex *P. aestuvarii*, BChla in vacuum (solid line) and excitation energies of BChla with surrounded MM charges (dashed line). DFT/MRCI, CAM-B3LYP, B3LYP and ZINDO/S taken from Refs. 126, 127.



**Figure 4.11** Corresponding shifts of excitation energies between vacuum and with MM charges. DFT/MRCI, CAM-B3LYP, B3LYP and ZINDO/S taken from Refs. 126, 127.

In Fig. 4.10, the excitation energies of the isolated BChl a chromophores are displayed in vacuum (solid lines) as well as the excitation energies after inclusion of the protein environment as point charges (dashed lines). Additionally, the BLA (in Å) values of each BChl a chromophore is plotted, which we calculated in the same way as described above. The BChl a chromophores differ in their BLA and shape. BChl a 1, 2, 3 and 7 display a small BLA, while BChl a 6 shows a higher BLA value. The hybrid functional B3LYP, the LC functional CAM-B3LYP as well as LC-DFTB and TD-DFTB lead to accurate vacuum excitation energies, which are in agreement with the DFT/MRCI method. Slightly deviations are obtained when the HF amount in the DFT functional is higher, as seen in the case of  $\omega$ B97X for the BChl a 5 and 6. A significantly different trend is given by SOS-ADC(2), which exhibits a behavior analogous to CC2, cf. Ref 127.

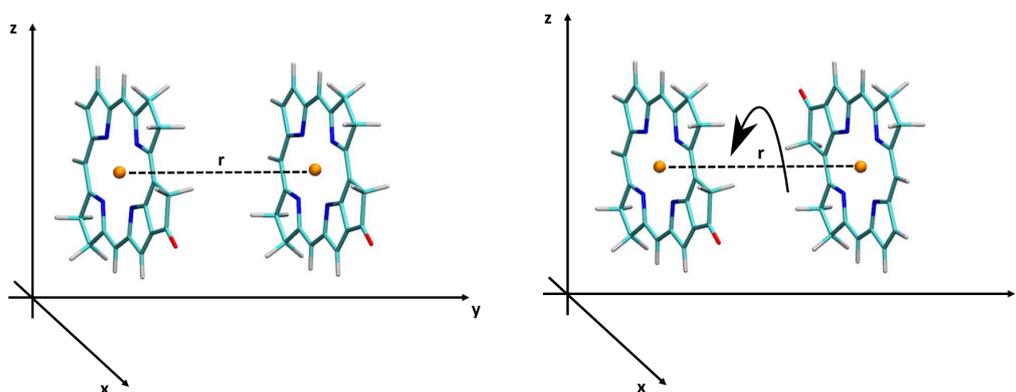
Fig. 4.11 displays the respective shift upon the protein environment for better comparison. In general, the reference DFT/MRCI method as already discussed above shows, that the BChl a chromophores exhibit only a small shift due to the electrostatic field in a range of about 0.0–0.1 eV. All methods except of SOS-ADC(2) can qualitatively reproduce the small shifts for the BChl a chromophores 1, 2, 3, 4 and 7. A slightly larger shift is obtained for BChl a 5 and the significant largest blue shift is received for BChl a 6, whose magnitude is dependent on the used method. An clearly overestimation is observed by ZINDO/S as previously reported, cf. Ref. 127. B3LYP slightly underestimates the shift, while TD-DFTB exhibits no shift. In the case of the LC functional, the HF amount displays a significant role,

i.e.  $\omega$ B97X obtains a higher shift than CAM-B3LYP. LC-DFTB exhibits also a blue shift and only slightly overestimates the shift when comparing to the reference method DFT/MRCI. The higher obtained shift in the case of BChl a might be due to the higher BLA value, however a detailed investigation is in the case of the FMO complex not given here, since sampled structures are necessary. Hence, these QM/MM optimized structures are not representative for an overall conclusions. Further studies on environmental effects are given for the LH2 complex in chapter 5.

In contrast to retinal, BChl a leads to only small shifts. The reason could be found in the different dipole moment changes upon excitation. The BChl a chromophore has a transition dipole moment of  $6\text{D}$ ,<sup>5,106</sup> while for the all-*trans* retinal with the protonated SB a transition dipole moment of about  $12\text{D}$ <sup>142</sup> is found.

### 4.4.3 Exciton Coupling

Due to the high amount of BChl a chromophores in LH complexes, exciton couplings occur and can be also responsible for shifting and broadening of the absorption spectra. The exciton couplings can be computed in two ways: (i) as a supermolecule, however restricted to a dimer or (ii) as Coulomb coupled monomers due to the spatially well separated BChl a chromophores. As mentioned above, the LH2 complex serves as an example how excitonic interaction can tune an absorption spectrum,<sup>43</sup> which is referred to as the B850 ring. The BChl a chromophores are narrower in the B850 ring than in the B800 ring.



**Figure 4.12** Artificial system built up of truncated BChl a dimers. Left: differ in distances  $r$ ; Right: differ in rotation.

Until now, there exist no systematic benchmark for exciton couplings in LH complexes. In order to benchmark LC-DFT as well as LC-DFTB for the computation of exciton couplings of BChl a chromophores, we set up an artificial system containing a truncated BChl a model, optimized it on the DFTB level of theory and build artificially dimers with different distances and rotation, see Fig. 4.12. This procedure avoids any geometrical influences on the exciton

coupling. The ADC(2) method is excluded from the benchmark study here due to the insufficient results in the previous section.

Supermolecule calculations were performed on the artificial dimers using CIS, TD-DFT with standard functionals as well as LC functionals, i.e. B3LYP,  $\omega$ B97X and CAM-B3LYP. Furthermore, supermolecule calculations are also performed with LC-DFTB and TD-DFTB. In Tab. 4.2 the results are shown, however not the resulted energy splitting, but the energy splittings divided by two. Since the exciton coupling benchmark is focused on the Coulomb couplings and in literature of LH complexes mostly the Coulomb couplings and not the energy splittings are presented, the same approach is also applied here.

**Table 4.2** The energy splitting obtained by supermolecule calculations are divided by two (in eV). The calculations are performed on artificially created dimers differing in the distance of a vacuum optimized truncated BChl a structure.

Distance ( $\text{\AA}$ )	CIS	B3LYP	TD-DFTB	$\omega$ B97X	CAM-B3LYP	LC-DFTB
7	0.069	0.051	0.031	0.059	0.058	0.063
8	0.051	0.039	0.024	0.044	0.044	0.048
9	0.040	0.030	0.019	0.034	0.034	0.037
10	0.031	0.023	0.015	0.027	0.026	0.030

**Table 4.3** The energy splitting obtained by supermolecule calculations are divided by two (in eV). The calculations are performed on artificially created dimers differing in the distance of a vacuum optimized truncated BChl a structure. The second molecule is rotated by  $180^\circ$ .

Distance ( $\text{\AA}$ )	CAM-B3LYP	LC-DFTB
7	0.056	0.065
8	0.043	0.050
9	0.033	0.035
10	0.026	0.030

As expected with an increasing distance between the dimers, the energy splittings become smaller. Comparing all used methods among each other only slight deviations are found except of TD-DFTB showing only half of the values with respect to the other methods. The large deviation by TD-DFTB results from the GGA functional on which it is based, since GGA functionals tend to underestimate the energy splittings and are also effected by state mixing.<sup>119</sup> CIS in contrast leads to slightly overestimated energy splittings when comparing to the DFT functionals, i.e. the hybrid and LC functionals. LC-DFTB is in good agreement to CAM-B3LYP and improves significantly compared to TD-DFTB.

In the LH2 complex, i.e. in the B850 ring, the BChl a chromophores are not parallel oriented but shifted to each other as well as rotated by about  $180^\circ$ , see Fig. 4.13. Hence, to account additional for the rotation, the second monomer of the dimer is rotated by  $180^\circ$ .

The results are presented in Tab. 4.3 using only CAM-B3LYP and LC-DFTB, due to the less variations obtained in Tab. 4.2 for all methods. The rotation has only negligible small effects, especially with a distance of 10 Å there is no difference obtained compared to the non rotating dimers.

The artificial dimers are taken to additionally compute the Coulomb couplings with LC-DFTB and TrESP without a scaling factor, i.e. which is normally applied to fit the transition dipole moments to experimental data. Concerning the explicit Coulomb couplings shown in Tab. 4.4, the TrESP method without the scaling factor can reproduce the respective supermolecule calculations. This was already reported by a previous study using TrESP based on B3LYP.<sup>101</sup> LC-DFTB in contrast overestimates the Coulomb couplings with respect to the supermolecule calculations by about +0.01 eV.

**Table 4.4** Coulomb couplings (in eV) of the artificial truncated BChl a dimer set using LC-DFTB and TrESP. The partial charges of the TrESP values are based on B3LYP and CAM-B3LYP.

Distance (Å)	LC-DFTB	TrESP*	
		B3LYP	CAM-B3LYP
7	0.082	0.050	0.056
8	0.063	0.038	0.043
9	0.049	0.029	0.033
10	0.039	0.023	0.026

\*taken by Ref. 141

The B850 ring of the LH2 complex (Fig. 4.13) is used to investigate the performance of LC-DFTB directly in the biological system. Firstly, supermolecule calculations are performed for every BChl a dimer using the LC-DFT functional CAM-B3LYP and LC-DFTB. Secondly, the Coulomb couplings are computed using LC-DFTB and TrESP. The BChl a dimers are subdivided into two groups denoted as  $\alpha - \beta$  dimer and  $\beta - \alpha$  dimer, see chapter 2.

The results shown in Tab. 4.5 for the biological systems reflects significantly differences compared to the artificial system. Dependent on the distance and rotation of the BChl a dimers the values are found in a range of about 0.04–0.08 eV. Mostly the values are larger than those obtained by the artificial system with respect to the distance.

The values obtained by supermolecule calculations of LC-DFTB are in good agreement to CAM-B3LYP. Even though the Coulomb couplings in the artificial system are overestimated by LC-DFTB, in the biological system it works qualitatively well, i.e. LC-DFTB can mostly reproduce the same trend than the supermolecule calculations. Besides that, the values are not always overestimated. For about half of the dimers the same values are obtained than with the supermolecule calculations. This reflects, that the orientation of the dimer to each other has a significant influence. TrESP shows also the same fluctuations of the values and differ slightly more to the supermolecule CAM-B3LYP calculation due to the

frozen TrESP approach. This approach was chosen since the same method would be applied on trajectories investigating Coulomb couplings in LH complexes. So that this approach is justifiable. Comparing LC-DFTB to TrESP, interestingly for one half of the dimers LC-DFTB yield the same range and fluctuation than TrESP. Hence, LC-DFTB yields quantitatively results either similar to TrESP or to the supermolecule calculations.

In this benchmark no scaling factor is applied. As introduced in section 4.1.3, previous computational studies applied mostly a scaling factor (see Tab. A.3 in the appendix), so that the comparison with these studies is here not possible. However, a recent study computed the Coulomb couplings with TDC using no scaling factor and obtained for the strongest Coulomb couplings values about 0.07 eV, which is in agreement to the results shown here.

**Table 4.5** The energy splitting obtained by supermolecule calculations are divided by two (in eV). The calculations are performed on BChl a dimers of the B850 ring of the LH2 complex. Distance measured from Mg to Mg ion. Coulomb couplings are given in eV.  $\beta - \alpha = a$  and  $\alpha - \beta = b$ .

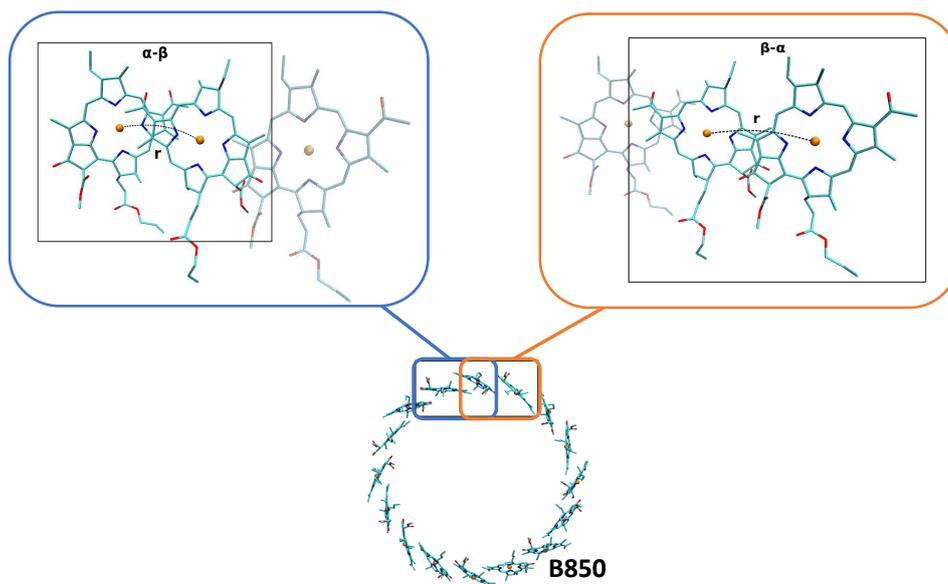
Distance (Å)	type	Supermolecule		Coulomb coupling	
		CAM-B3LYP	LC-DFTB	TrESP*	LC-DFTB
9.63	a	0.070	0.065	0.061	0.075
9.28	b	0.052	0.049	0.045	0.052
9.53	a	0.065	0.061	0.058	0.072
8.55	b	0.050	0.050	0.036	0.034
9.47	a	0.074	0.064	0.061	0.073
8.94	b	0.055	0.065	0.047	0.052
9.44	a	0.072	0.068	0.060	0.074
9.51	b	0.042	0.041	0.038	0.046
9.86	a	0.065	0.062	0.059	0.073
8.95	b	0.047	0.047	0.039	0.044
9.55	a	0.059	0.058	0.054	0.066
9.15	b	0.081	0.077	0.057	0.066
9.26	a	0.076	0.065	0.057	0.069
8.60	b	0.067	0.067	0.041	0.041
9.38	a	0.075	0.065	0.057	0.069
8.83	b	0.068	0.067	0.048	0.051

\*based on CAM-B3LYP and taken by Ref. 141

It also has to be mentioned, that the computational cost is especially in the case of this large biological systems an interesting point. Comparing the computational costs for the B850 ring, CAM-B3LYP with a parallel run on 6 cores takes about 2–3 days for one dimer. The LC-DFTB supermolecule calculations takes only about 15 minutes on a single core. And the Coulomb couplings computed with LC-DFTB are obtained in just few seconds. This

highlights the power of LC-DFTB and further benchmark are needed to investigate exciton couplings and especially the Coulomb couplings in LH complexes.

Furthermore, when applying TrESP for the computation of Coulomb couplings over a MD trajectory, the partial charges are frozen, since they are only determined for an equilibrium geometry and further on fixed used over the trajectory. LC-DFTB in contrast can compute the Coulomb couplings "on-the-fly" and is therefore a promising method with further improvements for the study of LH complexes.



**Figure 4.13** B850 ring of the LH2 complex. Accentuated are two BChl a dimers; Left:  $\alpha - \beta$  dimer and Right:  $\beta - \alpha$  dimer.

## 4.5 Conclusion

In this chapter, the recent time-dependent extension of the LC-DFTB method<sup>115</sup> is benchmarked on its application in biological systems, here on retinal and BChl a, the chromophores of rhodopsins and LH complexes. This systematic benchmark uses various methods and focuses on several influencing factors responsible for color-tuning.

Concerning the chromophore retinal, LC-DFT functionals as well as LC-DFTB show an improved behavior to describe the geometrical impact on excitation energies with respect to the BLA as well as the twist of the  $\beta$ -ionone ring. The influence of the electrostatic environment is better described by the LC-DFT functional and LC-DFTB than by GGA or hybrid DFT functionals, however showing still a color-weakness. In comparison of the three LC-DFT functionals,  $\omega$ B97X and LC-BLYP lead to better results because of the higher HF-X amount in the long-range limit, however with errors in the absolute excitation energies.

In the case of BChl a, LC functionals and LC-DFTB yield the same trend than ADC(2) to describe the excitation energies with regard to various BChl a geometries. Within this study DFTB leads to BChl a geometries with the same accuracy than B3LYP, which offers therefore a good optimization method for further QM/MM optimization or MD simulations to yield an accurate description of the chromophore ground state geometry. The inclusion of the protein environment reveals, that LC-DFTB can reproduce the same trend as DFT/MRCI and even better than CAM-B3LYP. In contrast to ZINDO/S, LC-DFTB does not overestimate the excitation energies by including the electrostatic environment. The influence of the protein environment on the excitation energies of BChl a chromophores leads in general to only small shifts, in contrast to retinal, where the protein environment has a significant contribution on the absorption maxima shifts.

The benchmark study on the exciton coupling performing supermolecule as well as Coulomb coupling calculations on an artificial system, i.e. a truncated BChl a structure optimized in vacuum, shows that LC-DFTB supermolecule obtains the same results as LC-DFT functionals. The Coulomb couplings using LC-DFTB are however slightly overestimated compared to the supermolecule calculations. The further application on the biological system, i.e. the B850 ring system of LH2, reveals however that LC-DFTB can nicely reproduce qualitatively the same trend as the supermolecule calculations using LC-DFT. Quantitatively not an overall overestimation is observed, so that further benchmark are necessary to figure out, if for example a simple scaling factor is necessary to yield the accuracy of the supermolecule calculations.

In general, LC-DFTB shows that it can reproduce the same trend as LC-DFT functionals describing various influencing factors responsible for color tuning. Apart from the weaknesses in the case of retinal, it has a big potential for further developments to study such chromophores. Further benchmarks concerning the geometry optimization with LC-DFTB will also give more insights about its performance. And the implementation of gradients for the excited state might provide a useful method for study excited state pathways of retinal. In the case of the LH complexes, LC-DFTB even show a good behavior with respect to the influences of the protein environment on the excitation energies. Even though LC-DFTB describes some weaknesses in the description of Coulomb couplings, the efficient calculation of TD-DFTB couplings within mulliken charges was already pointed out, cf. Ref. 179. Furthermore, the charges can be obtained "on-the-fly" in DFTB QM/MM MD simulations, whereas for e.g. TrESP this is not possible, since the computation is based on fixed transition charges projected on the trajectory. LC-DFTB has low computational costs, being about three orders faster than full TD-DFT with a minimum of loss in accuracy.<sup>115</sup> Hence, LC-DFTB is promising for the study of LH complexes, for example to study the dynamic exciton energy transfer towards the reaction center.



---

## Performance of LC-DFTB on Rhodopsins and Light-Harvesting Complex II

### 5.1 Introduction

The benchmark study in chapter 4 of LC-DFTB on retinal and bacteriochlorophyll (BChl), the chromophores of rhodopsins and light-harvesting (LH) complexes, reveals how necessary it is to perform a systematic benchmark using various methods in order to investigate all influencing factors responsible for color tuning. The investigation of a single chromophore structure does however not account for the full shape of the absorption spectrum. The width or possible shoulders of an absorption spectrum are also unique for the respective biological system.

QM/MM optimized models can take the influence of the protein environment on the chromophore already into account. For example, rhodopsins like bacteriorhodopsin (bR) or pharaonis phoborhodopsin (ppR) exhibit a strong hydrogen bond network in the active site confining the chromophore retinal in its structural fluctuation. Hence, a QM/MM optimized model could reveal even a well defined minimum. To account also for several structural and environmental fluctuations, molecular dynamic (MD) simulations have to be performed. MD simulations describe for those proteins only an oscillation of the chromophore retinal around the energy minimum.<sup>82</sup> However, a single QM/MM optimized structure may not necessarily account for the absorption maximum peak, as shown for another rhodopsin, the channelrhodopsin-2 (ChR 2) exhibiting a multi-peak absorption spectrum.<sup>73</sup> In these cases sampling of excitation energies is important, since there might be a not well defined global minimum within  $k_B T$  and thus the absorption spectra may sample several local minima of the active site structure. Furthermore, the absorption maximum or mean absorption energy can be significantly shifted compared to the optimized geometry, i.e. the mean structure deviates from the optimized one. Temperature can also have an effect on the absorption spectrum as shown e.g. for bR<sup>233</sup> and the LH2 complex.<sup>42</sup>

In general, the broadening of the spectrum results from fluctuations of the chromophore, as well as from fluctuations of the protein environment. Fluctuations of the chromophore structure manifest themselves as fluctuations of the BLA, since the C-C stretch mode plays a dominant role. At this point, the proper description of excitation energies vs. geometry changes becomes important. Furthermore, the fluctuations of the environment are reflected in fluctuations of the external Coulomb potential, and therefore the exact response of the excitation energy to this external perturbation is a second important property that must be described correctly.

In this chapter, the benchmark study of chapter 4 is extended, i.e. with the performance of LC-TD-DFTB on rhodopsins and LH complexes. In rhodopsins it is well known that QM/MM MD simulations are needed to correctly describe the active site and the related hydrogen bond network.<sup>49,225</sup> In the case of the LH2 complex however, there exist no QM/MM MD study especially performed with DFTB3, but highly desirable by literature.<sup>97</sup> Hence, QM/MM MD simulations of the LH2 complex are performed and compared to the sampled structures obtained by MM MD simulations with respect to i.a. the bond length alternation. MM MD simulations are used by default for the study of LH complexes due to high amount of chromophores and the large size of the system as a whole.

## 5.2 Computational Details

### 5.2.1 Models

The rhodopsin model systems used in section 5.3, bacteriorhodopsin (bR) (PDB code 1C3W) and phoborhodopsin (ppR) (PDB code 1H68), were taken from Ref. 59, 82, 222. Both models contain one monomer solvated in water with the TIP3P<sup>95</sup> water model. For the study of the LH complexes, the light-harvesting complex II (LH2) (PDB code 1LGH)<sup>109</sup> from *Rhodospirillum molischanum* was taken. The system setup was provided by the Kleinekathöfer group.<sup>105</sup> The protein is embedded in a POPC (1-Palmitoyl-2-oleoylphosphatidylcholine) lipid bilayer with about 30 Å of water (TIP3P) on both sides and 16 Cl<sup>-</sup> ions for neutralization.

### 5.2.2 QM/MM MD Simulation

QM/MM MD simulations were performed for both rhodopsins and the LH2 complex. For rhodopsins it is well known, that especially the active site of the protein has to be described with QM methods to gain a correct description of the hydrogen bond network.<sup>49,225,233</sup> A well chosen QM region is important for retinal proteins to consider effects, like charge transfer, protein polarization and dispersion interaction.<sup>220</sup>

The QM region of bacteriorhodopsin (bR) contains the chromophore retinal, the two counterions D85 and D212, three water molecules near the RSBH+ and R82 hydrogen bonded to one QM water. Analogous to bR, the same QM region was taken for ppR, i.e. the two

counterions D75 and D201 as well as R72. Both active sites containing the involved QM region are shown in chapter 2, Fig. 2.8. The remaining part of the system was treated with MM using the CHARMM36 force field.<sup>84</sup> For the QM part the DFTB3/3OB method (extended self-consistent-charge Density-Functional Tight-Binding)<sup>63–65</sup> is used, implemented in the GROMACS package.<sup>118</sup> DFTB3/3OB is advantageous for describing a reliable model of hydrogen bonded structures as present in rhodopsins, since it is able to describe hydrogen-bonded networks with similar accuracy as full DFT calculations performed with medium sized basis sets.<sup>63</sup> The QM/MM MD trajectories have a length of 1 ns, respectively.

In the case of the LH2 complex, QM/MM optimization as well as QM/MM MD simulations were performed. The QM/MM optimizations were performed using the DFTB3/3OB method for the QM region as implemented in the GROMACS package (version 2017.1).<sup>1,118</sup> The remaining part of the system was treated with MM using the CHARMM27 force field.<sup>58,138</sup>

The QM region of the LH2 complex contains one BChl a chromophore, respectively. The phytyl tail was not included in the QM region, since the QM/MM bond is defined between the CP1 and CP2 atoms. The neglect of the phytyl-tail in the QM region saves computational cost and does not effect significantly the excitation energies, cf. Ref. 39,125. The system was QM/MM minimized using the steepest descent (SD) algorithm (100000 steps with a tolerance of  $1000 \text{ kJ mol}^{-1} \text{ nm}^{-1}$ ).

The QM/MM MD simulations were performed analogously with the DFTB3/3OB method for the QM region as implemented in the GROMACS package (version 2017.1),<sup>1,118</sup> while the remaining part of the system was treated with MM using the CHARMM27 force field.<sup>58,138</sup> The QM region was the same as described above for the QM/MM optimization. The QM/MM MD trajectories have a time length of 1 ns, respectively.

### 5.2.3 MM MD Simulation

In the case of the LH2 complex also MM MD simulations were performed. The MM MD simulation was carried out with the GROMACS package (version 2016.3)<sup>80</sup> and the CHARMM27 force field.<sup>58,138</sup> Firstly, the LH2 complex was minimized using the steepest descent algorithm (50000 steps with a tolerance of  $1000 \text{ kJ mol}^{-1} \text{ nm}^{-1}$ ). Subsequently, the system was equilibrated with a time length of 3 ns and a temperature of 300 K using the Nosé-Hoover thermostat<sup>83,168</sup> (time step 1 fs). The heavy atoms of the protein and all chromophores were restrained to their initial positions by harmonic potentials (force constant:  $1000 \text{ kJ mol}^{-1} \text{ nm}^{-1}$ ). Additionally, an NPT equilibration was performed with a time length of 8 ns and a pressure of 1 bar using the Parrinello-Rahman barostat<sup>173</sup> (time step 2 fs). Subsequently, the restraints were removed and an NPT equilibration of 1 ns was performed. After the equilibration, a MM MD simulation was performed with a time length of 1 ns. During the simulation a periodic boundary condition was applied and the long-range interactions were considered by means of the Particle-Mesh Ewald method.

### 5.2.4 Excitation Energies

The QM/MM optimized structures of bR and ppR taken from Ref. 82 were used to compute the excitation energies using several methods, analogous to the benchmark of LC-DFTB described in chapter 4. For the wavefunction methods CC2, SOS-CC2, ADC(2) and SOS-ADC(2) were chosen and all calculations were performed by S. Höfener using the Turbomole program package.<sup>2</sup> For the TD-DFT calculations, the LC functionals  $\omega$ B97X and CAM-B3LYP were chosen. Furthermore, the excitation energies were also computed with LC-TD-DFTB (further on short-hand denoted as LC-DFTB) and TD-DFTB. All details can be found in chapter 4, section 4.2.3. The excitation energy calculations were performed in vacuum, i.e. only a single chromophore and with the inclusion of the protein environment, i.e the remaining part of the system.

The sampled structures of the QM/MM MD simulations of both rhodopsins were taken for subsequent excitation energy computation using LC-DFTB and OM2/MRCI. The OM2/MRCI calculations were carried out using the MNDO2005 program package.<sup>113,226</sup> An active space of (20, 20) was used for the simulated structures of the rhodopsins.<sup>73</sup> For each 1 ns long QM/MM MD trajectory, the excitation energies of 1000 snapshots were computed, respectively.

In the case of the LH2 complex, the QM/MM optimized structures were also taken for the computation of excitation energies using LC-DFTB and ZINDO/S-CIS. ZINDO/S-CIS is implemented in the ORCA program package<sup>158</sup> and denoted ZINDO/S in the following. For the ZINDO/S calculations an active space of (10,10), i.e. 10 electrons in 10 spatial orbitals, was used, which yields a sufficient agreement with experimental results.<sup>131,172</sup>

Additionally, the sampled structures from MM and QM/MM MD simulations were used for subsequent computation of excitation energies using LC-DFTB and ZINDO/S with the same conditions as described before. For each 1 ns long MM as well as QM/MM MD simulations, the excitation energies of 1000 snapshots were computed, respectively.

The computation of the excitation energies were performed in vacuum, i.e. the single BChl a chromophore and with inclusion of the protein environment. The phytyl tail was also not considered in the QM region for the excitation energy computation. For the protein environment a charge sphere around the Mg atom of every BChl a chromophore with a size of 20 Å was chosen as MM environment. The atoms within a radius of 18 Å were considered with full MM point charges, whereas for the atoms within the radius  $r$  with  $18 \text{ \AA} \leq r \leq 20 \text{ \AA}$  a switch function was applied to avoid discontinuity. The switch function described a linear declining function from one to zero. In general, 48 000 excitation energy calculations were performed for the LH2 complex (24 000 vacuum and 24 000 with MM environment).

### 5.2.5 Coulomb Coupling

The Coulomb couplings using LC-DFTB were computed for the QM/MM optimized structures as well as for the sampled structures of the MM and the QM/MM MD simulations.

Firstly, only the Coulomb couplings between neighboring BChl a chromophores were calculated, resulted in  $2 \times 2$  matrices. In this matrices, the diagonal elements are given by the excitation energies of the monomers and the off-diagonal matrix elements are given by the coupling matrix elements. The diagonalization of the matrices gives the eigenvalues. Secondly, the Coulomb couplings over the whole ring were calculated, respectively. This results in a  $8 \times 8$  for the B800 ring and a  $16 \times 16$  matrix for the B850 ring.

The lowest eigenvalues of the matrices are used, since the supermolecule calculations performed in the previous chapter show, that the oscillator strength for the lower eigenvalue is nearly one and an only small oscillator strength for the upper eigenvalue is found.

## Results and Discussion

### 5.3 Rhodopsins

To test the performance of LC DFT functionals and LC-DFTB on rhodopsins, bR and ppR serve as rhodopsin models. Both rhodopsins share the same basic structure, i.e. they are build up by seven transmembrane  $\alpha$ -helices embedded in a membrane and forming an internal binding pocket, where the chromophore all-*trans* retinal is bonded via a Schiff base to an  $\epsilon$ -amino group of a lysine side chain of helix G (helix 7). The binding pocket of bR and ppR differs in only 10 amino acids within  $5 \text{ \AA}$  of the retinal. As introduced in chapter 2, the active site of bR and ppR contains two counterions forming a typical pentagonal hydrogen bond cluster with three water molecules near the SB. Besides the similarities, the absorption maximum of ppR ( $497 \text{ nm}$ )<sup>38</sup> is blue shifted by about  $70 \text{ nm}$  relative to bR ( $568 \text{ nm}$ ).<sup>24</sup> The computation of this shift shows clearly the performance of the method since individual protein shifts might be accompanied with error cancellation.

As introduced in chapter 4, the color-tuning in rhodopsins is effected by different influences. This large blue shift of ppR with respect to bR has therefore several reasons. The two main and equally contributed reasons are the different neutral amino acids in the binding pocket of retinal and the difference in the extended hydrogen bond network in the extracellular side involving the counterion complex of the protein. Together these are responsible for about 90 % of the shift. The differences in the retinal conformations is only small, while ppR has a higher BLA than bR, induced by the electrostatic environment. The contribution to the shift is therefore small and takes about 10 %.<sup>82, 220</sup>

These differences lead to a differently polarized retinal chromophore by the polar and charged groups of bR and ppR resulting in a change of the dipole moment compared to the vacuum and in the BLA along the polyene chain. Both changes are larger in ppR than in bR, which indicates a stronger electrostatic interaction with the protein in ppR.<sup>82, 220</sup>

### 5.3.1 QM/MM Optimized Models

In Tab. 5.1, the excitation energies of retinal in different electrostatic environments are compared. For each protein, the QM/MM optimized structures of bR and ppR from Refs. 59, 82, 222 are taken. The excitation energies are computed in vacuum (omitting the MM point charges) and within the protein environment by different QM methods.

All methods agree that without point charges, the excitation energies in bR and ppR are very similar and differ by only 0.03 eV in average. This is due to the similar geometry of retinal in the two proteins. Also when retinal is optimized in vacuum, excitation energies change very little as described in chapter 4, see Tab. A.1 in the appendix. When including the protein environment, ppR gives a larger blue shift than bR. All methods reproduce this trend, but the magnitude of the shifts varies and reflects the accuracy of the methods in describing the electrostatic influence of the environment. When comparing with the experimental shifts of 0.18 and 0.50 eV for bR and ppR, respectively, it appears that all wavefunction based methods overestimate the shift w.r.t. vacuum, whereas DFT based methods underestimate only the shift between the two proteins.

Previous works, however, have shown that point charges obtained by electronic embedding overpolarize the chromophore region, which is improved by polarizable embedding. When replacing the point charges in the binding pocket by the electron density of a separate QM region,<sup>221</sup> or by a polarizable force field,<sup>219</sup> significantly lower excitation energies are obtained. When the chromophore is described with SORCI and the rest of the protein with the polar.h<sup>219,221</sup> polarisation model, excitation energies of 2.16 and 2.42 eV have been reported for bR and ppR, respectively.<sup>219,221</sup> Compared to vacuum (2.00 eV),<sup>167</sup> this corresponds to shifts of 0.16 and 0.42 eV, which are in good agreement with the experiment. For this reason, the discussion is continued using the SORCI shifts as a reference for the other methods.

The respective shifts obtained with OM2/MRCI, CC2, and ADC(2) are in the range of the SORCI ones, with an average deviation of 0.02–0.04 eV. The SOS variants of CC2 and ADC(2) yield shifts that are larger by 0.06 eV on average. The LC functionals  $\omega$ B97X and CAM-B3LYP underestimate the shifts by 0.20 and 0.30 eV in average, respectively. TD-DFTB underestimates the shift by even up to 0.35 eV with respect to SORCI. Surprisingly, LC-DFTB does not improve the shifts, neither.

These trends are very similar to the ones shown in the previous chapter for the model chromophore in presence of a single point charge and consistent with previous works (see Ref. 222 and references therein), which show that also CIS (based on HF or an OM2 Hamiltonian) underestimates the shift in bR. Therefore, it is not surprising that the shift obtained with LC functionals is larger than that of pure or fixed-hybrid DFT functionals but still significantly lower than that of higher-level methods, like SORCI or CASPT2. Regarding the shifts, the  $\omega$ B97X functional performs better than CAM-B3LYP due to its larger amount of HF-X in the long-range limit, but this improvement is paid with a larger error in absolute excitation

energies.

The presented results for bR show that errors implicit to the QM/MM approach can result in fortuitous error cancellation when considering an individual shift but not when predicting properties of several proteins and/or mutants. Apart from the electrostatic representation of the MM region, charge transfer between the QM and the MM regions is suppressed, which can vary, for instance, with the strength of hydrogen bonds or distance to the counter ion and can lead to significant errors.<sup>82,191,222</sup> The description of such inter-residual charge transfer might also be improved by LC functionals as it is known that local DFT functionals tend to delocalize electronic states too much due to the self-interaction error.<sup>112</sup>

**Table 5.1** Lowest excitation energies and shifts of QM/MM optimized structures of bR and pPR, all in eV. (pPR - bR) denotes the difference of the two systems including the protein environment. The excitation energies are computed only with static external charges without polarization effects leading to higher excitation energies. Results from CC2, SOS-CC2, ADC(2) and SOS-ADC(2) are taken from Ref. 86.

	Exp.	Wavefunction						DFT		DFTB	
		SORCI	OM2/MRCI	CC2	SOS-CC2	ADC(2)	SOS-ADC(2)	$\omega$ B97X	CAM-B3LYP	LC-DFTB	TD-DFTB
<b>bR</b>											
vacuum	2.00 <sup>a</sup>	1.86 <sup>d</sup>	2.22 <sup>d</sup>	2.08	2.25	1.89	2.13	2.65	2.64	2.58	2.04
protein	2.18 <sup>b</sup>	2.34 <sup>d</sup>	2.66 <sup>d</sup>	2.48	2.76	2.35	2.64	2.94	2.85	2.77	2.22
shift	+0.18	+0.48	+0.44	+0.40	+0.51	+0.46	+0.51	+0.29	+0.21	+0.19	+0.18
<b>pPR</b>											
vacuum	2.00 <sup>a</sup>	1.91 <sup>d</sup>	2.24 <sup>d</sup>	2.08	2.30	1.92	2.20	2.67	2.65	2.60	2.00
protein	2.50 <sup>c</sup>	2.63 <sup>d</sup>	2.96 <sup>d</sup>	2.80	3.11	2.70	3.01	3.18	3.06	2.95	2.37
shift	+0.50	+0.72	+0.72	+0.72	+0.81	+0.78	+0.81	+0.51	+0.39	+0.35	+0.37
pPR - bR	+0.32	+0.29	+0.30	+0.32	+0.35	+0.35	+0.37	+0.24	+0.21	+0.18	+0.16

<sup>a</sup> Ref. 167 <sup>b</sup> Ref. 24 <sup>c</sup> Ref. 38 <sup>d</sup> Ref. 82

### 5.3.2 QM/MM MD Simulation

A static structure of retinal with its surrounded protein environment obtained by QM/MM optimization misses the fluctuation of the protein environment and the retinal conformations. A QM/MM MD simulation is carried out to sample several structures to assess other influences such as temperature, which are subsequently used for the computation of excitation energies. Thus, the excitation energies are computed on 1000 snapshots using LC-DFTB and OM2/MRCI for bR and ppR, respectively. All excitation energies are weighted by the oscillator strength and plotted as histograms, see Fig. 5.1. The absorption maxima are obtained with a fit using a Gaussian function and displayed in Tab. 5.2.

**Table 5.2** Absorption maxima (eV) of bR and ppR, obtained by computation of excitation energies on sampled structures by QM/MM MD simulation with subsequent Gaussian fit of the histograms. (ppR - bR) denotes the difference of the two systems in the protein environment.

	Exp.	OM2/MRCI	LC-DFTB
bR			
vacuum	-	2.24	2.52
protein	2.18 <sup>a</sup>	2.65	2.73
shift	-	+0.41	+0.21
ppR			
vacuum	-	2.26	2.53
protein	2.50 <sup>b</sup>	2.98	2.93
shift	-	+0.72	+0.40
ppR - bR	+0.32	+0.33	+0.20

<sup>a</sup> Ref. 24 <sup>b</sup> Ref. 38

The absorption maxima of the retinal chromophore in bR and ppR in vacuum differ of about  $\pm 0.02$  eV for each method, respectively. This is expected, since the retinal conformations are similar in bR and ppR. In comparison with the QM/MM optimized models, OM2/MRCI displays a slight blue shift of +0.02 eV and LC-DFTB a red shift of about -0.06 eV for the vacuum excitation energies. The same behavior is obtained for the absorption maxima including the protein environment.

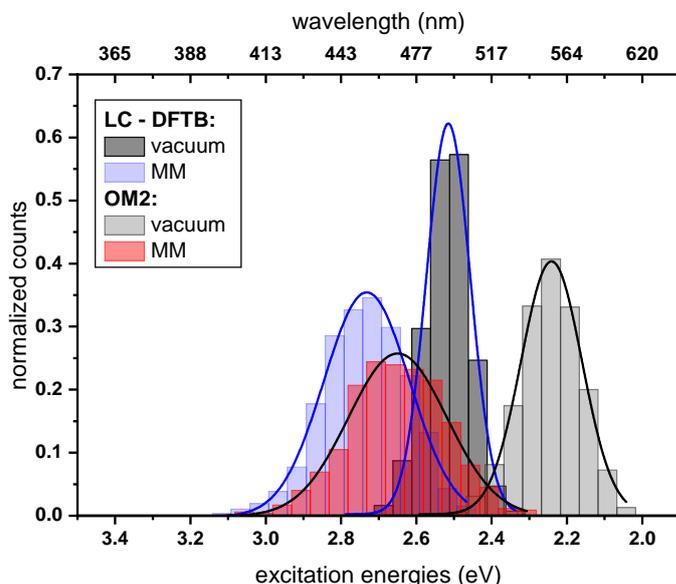
In bR and ppR the inclusion of the protein environment results for both methods in a blue shift. In the case of bR, OM2/MRCI leads to a slightly smaller blue shift (0.41 eV) compared to the QM/MM optimized models (0.44 eV). Whereas in the case of ppR, LC-DFTB computes a slightly higher blue shift (0.40 eV) compared to the QM/MM optimized models (0.35 eV).

The slight deviations compared to the QM/MM optimized models of both rhodopsins bR and ppR is mainly reflected by the strong hydrogen bond network in the active site confining the chromophore retinal in its structural fluctuation. In this case the QM/MM

MD simulations describe only an oscillation of the chromophore retinal around the energy minimum.<sup>82</sup> However, it has to be mentioned that rhodopsins exist, e.g. ChR2-WT exhibiting a very flexible structure and needs therefore sampling by QM/MM MD simulations.<sup>73</sup>

The difference of the absorption maxima between both rhodopsins (ppR-bR) obtained by OM2/MRCI (+0.33 eV) is as expected in agreement to the experimental shift (+0.32 eV). LC-DFTB shows with +0.20 eV a smaller shift of the absorption maxima between bR and ppR. Here, the color-weakness of LC-DFTB is also visible.

The computed excitation energies of the sampled structures are plotted as histograms in Fig. 5.1, here for bR. The absorption spectra for ppR are given in the appendix, see Fig. A.2. In the case of OM2/MRCI, a broad simulated absorption spectrum is obtained, while LC-DFTB tends to yield less broad absorption spectra, especially when considering only the retinal chromophore without the protein environment. The sampled retinal conformations differ in the BLA values and leads therefore to different excitation energies, as described in the previous chapter. Additionally, also the protein environment fluctuation is taken into account, which could also have influences since the shift is for example dependent on the retinal chromophore and counterions distances. Hence, weaknesses of LC-DFTB becomes visible, since it can not reproduce the same broadness as OM2/MRCI.



**Figure 5.1** Simulated absorption spectrum of bR. LC-DFTB and OM2/MRCI are used for the computation of the excitation energies. The histograms are based on snapshot geometries of QM/MM MD trajectories (1 ns length). Plotted are the excitation energies weighted by the oscillator strength for (i) only the retinal chromophore (vacuum) and (ii) with additional fixed MM point charges to account for the protein environment (MM). Gaussian functions are used to determine the corresponding maxima, in blue: LC-DFTB and in black: OM2/MRCI.

## 5.4 Light-Harvesting Complex II

The light-harvesting complex II (LH2) of purple bacteria from *Rhodospirillum rubrum* as introduced in chapter 2 is one of the most investigated LH complex due to its closely symmetric arranged BChl a chromophores in the B850 ring. Besides the BChl a chromophores, the LH2 complex consists also of carotenoids, which form a subunit with three BChl a chromophores as shown in chapter 2 (Fig. 2.13). Here, we consider only the BChl a chromophores.

The LH2 complex exhibits a characteristic absorption spectrum<sup>66</sup> with two significant maxima found at around 800 nm (1.55 eV) and 850 nm (1.46 eV) and denoted as the B800 and B850 ring. The B800 rings contains 8 BChl a chromophores, which are separated by a distance of about 20 Å, measured from the Mg to Mg ion. This ring is oriented parallel to the membrane plane and perpendicular to the other BChl a ring, the B850 ring. The B850 ring contains 16 BChl a chromophores, which are more compactly arranged.

The LH2 complex serves therefore as an example how excitonic interaction can tune an absorption spectrum,<sup>43</sup> which is thus referred to as the B850 ring. The absorption maxima of the B800 ring is set to be mainly influenced by the protein environment, since it is near to the BChl monomer absorption maxima, which is about 770 nm (1.61 eV) in organic solvents.<sup>213</sup>

The LH2 complex serves here as a model to test on the one hand the performance of LC-DFTB to investigate such proteins. On the other hand, QM/MM MD simulations are performed with the LH2 complexes in order to compare the description by QM/MM with respect to the MM MD simulations.

### 5.4.1 QM/MM Optimized Model

#### Environmental Effects

The influence of the protein environment on the excitation energies of the BChl a chromophores in the LH2 complexes are investigated on QM/MM optimized models using LC-DFTB and ZINDO/S. Tab. 5.3 shows the excitation energies of BChl a chromophores in vacuum, i.e. only the single chromophores are taken into account as well as the excitation energies by including the protein environment as fixed MM point charges. The resulted shifts are also displayed. The excitation energies are averaged for all BChl a chromophores over the ring, respectively.

Both rings, B800 and B850, show no differences in the vacuum excitation energies for both methods LC-DFTB and ZINDO/S, since the BLA values are similar for both rings. The BLA in B800 is around 0.01 Å, while in B850 it tends to be  $\leq 0.01$  Å. In the appendix, the BLA values of the QM/MM optimized BChl a chromophores can be found in Tab. A.8. When DFTB is used for the optimization of one BChl a chromophore in vacuum, a BLA value of 0.004 Å is found, see chapter 4, Fig. 4.7. The BLA values especially in the B800 ring are slightly higher than the vacuum optimized ones and reflects changes by the protein environment.

**Table 5.3** Excitation energies of BChl a monomers of LH2 in vacuum and with protein environment using LC-DFTB and ZINDO/S (in eV). The excitation energies are averaged over the B800 and B850 ring, respectively.

	LC-DFTB			ZINDO/S		
	vacuum	protein	shift	vacuum	protein	shift
B800	1.83	1.84	+0.01	1.46	1.50	+0.04
B850	1.83	1.83	$\pm 0.00$	1.46	1.46	$\pm 0.00$

The protein environment changes the shape of the BChl a geometry. While the DFTB vacuum optimized BChl a geometry is planar, in mostly all B800 BChl a chromophores a bowl shape is found. This was suggested to be effected by the out-of-plane displacement of the Mg ion due to the axial ligand coordinated to the central Mg ion "pulling" it out of the plane.<sup>212</sup> The influences of the axial ligand or other hydrogen bonded amino acids to the BChl a chromophore seems however to have few impact on the excitation energy, cf. e.g. Ref. 9, 127, 159.

A previous study investigated the influence of neighboring residues of BChl a on the excitation energy in LH2 from *Rhodospseudomonas acidophila* (27 BChl a) using RASPT2.<sup>9</sup> Taking the nearby residues in the QM part into account, the axial residues lead to a red shift of 16-45 meV, while hydrogen bonded residues lead to a blue shift of 3-35 meV. The impact of the influence of neighbored residues seems to be small and these values belonging to a few meV range below the methods errors, making it difficult to estimate the accuracy of the resulted shifts.

This is also shown here, the inclusion of the protein environment leads to slightly higher shifts in the B800 ring than in the B850 ring with respect to the individual BChl a chromophores, see appendix, Tab. A.8. The averaged excitation energies show even a slightly higher shift of +0.04 eV for the B800 ring obtained by ZINDO/S. LC-DFTB give here improved results compared to ZINDO/S, since the benchmark study in chapter 4 showed that LC-DFTB did not overestimate the excitation energies when considering the environmental effects. ZINDO/S in contrast obtained an overestimation.

Hence, the protein environment seems not effect much changes compared to the single BChl a chromophore in vacuum. The shifts of absorption maxima should be effected more by the exciton coupling of the several BChl a chromophores and is discussed in the next section.

### Exciton Coupling

The Coulomb couplings are computed of the QM/MM optimized model using LC-DFTB. Firstly, the Coulomb couplings are calculated for the BChl a of neighboring chromophores showing strong couplings. This results in dimer Hamiltonians, which diagonalizations obtain the eigenvalues. The lowest eigenvalues of each dimer are averaged over both rings, respectively. Secondly, the Coulomb couplings of BChl a dimers are calculated with all possible combinations

during one ring, resulted in one 8x8 matrix for the B800 ring and one 16x16 matrix for the B850 ring. The lowest resulted eigenvalues of both rings is here also taken. The results are displayed in Tab. 5.4. The exciton couplings of B800-B850 inter ring are neglected here, since these might be weak, but should be in the same range as the Coulomb couplings in the B800 ring due to similar distances.<sup>190</sup>

**Table 5.4** Lowest BChl a dimer energy of LH2 resulted by diagonalization of the dimer exciton hamiltonian, because of the highest oscillator strength. The energies are averaged over the B800 and B850 ring, respectively. Additionally, lowest energy resulted by diagonalization of the (8x8) and (16x16) matrix for B800 and B850, without considering the oscillator strength. All excitation energies and Coulomb couplings include the protein environment and are calculated using LC-DFTB (in eV).

		B800	B850	B850 shift
LC-DFTB	monomor	1.84	1.83	-0.01
	dimer	1.82	1.77	-0.05
	complete ring	1.79	1.73	-0.06
Exp.		1.55	1.46	-0.09

The neighboring BChl a dimers in the B800 ring show Coulomb couplings of about 0.01 eV, while the other intra BChl a dimers lead to values smaller than 0.01 eV. In B850, the neighboring BChl a dimers display Coulomb couplings in a range of around 0.08 eV to 0.03 eV depending of the dimer type, i.e.  $\alpha - \beta$  or  $\beta - \alpha$  dimer. Negligible small shifts are also observed for the other intra BChl a dimers. The Coulomb couplings of B800 of neighboring BChl a chromophores are in the same range as obtained by previous studies.<sup>42,172</sup> However, the Coulomb couplings of the B850 ring are overestimated, when comparing to previous studies using e.g. TrESP or TDC obtaining Coulomb couplings of about 0.03–0.04 eV.<sup>42,117,139</sup> However, these methods applied a scaling factor to fit to experimental data. When no scaling factor was applied, as the case in a recent study using TDC, Coulomb couplings of about 0.07 eV are reported,<sup>192</sup> similar to the LC-DFTB values.

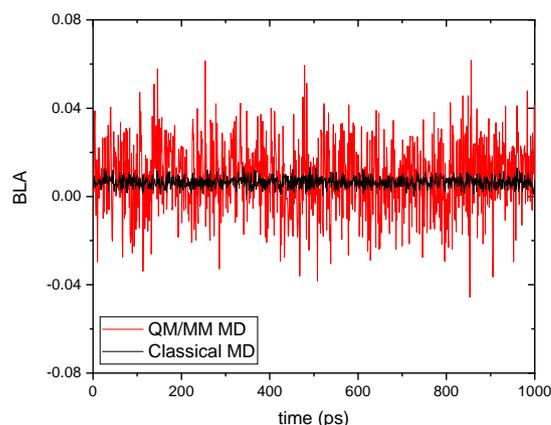
The inclusion of the exciton couplings leads for both rings to a red shift of the absorption maxima. Considering only the protein environment does not explain the experimental shift between the absorption maxima of both rings. However, ZINDO/S already achieve a shift of around 40% of the experimental shift, which originates from the overestimation of the protein environment influence. Despite the overestimation of the Coulomb couplings, only a shift of about 70% of the experimental shift is observed when considering all excitonic couplings. However, here LC-DFTB is applied on only QM/MM optimized models, while sampling might lead to a different picture.

### 5.4.2 QM/MM and MM MD Simulation

#### Bond Length Alternation

The sampled structures obtained by MM as well as QM/MM MD simulations are used to investigate BLA changes. In chapter 4, the computation of the BLA values of BChl a chromophores optimized in vacuum by different methods reflects, that an only low BLA value (around  $0.004 \text{ \AA}$ ) describes correctly the BChl a conformation. HF for example obtains a value of around  $0.102 \text{ \AA}$  and thus fails to describe the BChl a geometries correctly. As described in the previous section 5.4.1, the protein environment changes also the BLA, however with minor effects.

The comparison of the BLA values over the MM and the QM/MM MD trajectory shows, that the QM/MM obtained BLA values obtain a complete different picture, as displayed in Fig. 5.2. The BLA values of all BChl a chromophores of every ring are computed. In Fig. 5.2 only one BChl a chromophore of the B800 ring is chosen as example, since every BChl a chromophore show the same behavior. The following analysis is related to this BChl a chromophore, which represents all other BChl a chromophores.

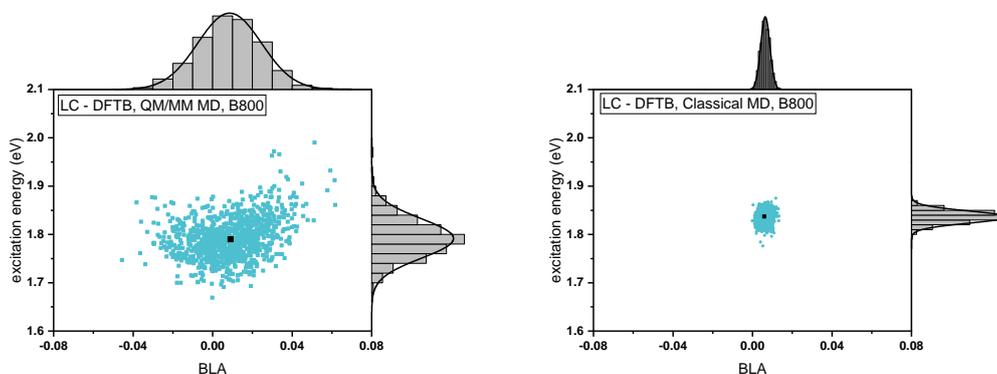


**Figure 5.2** BLA (in  $\text{\AA}$ ) of one BChl a of the B800 ring plotted over MM and QM/MM MD trajectories.

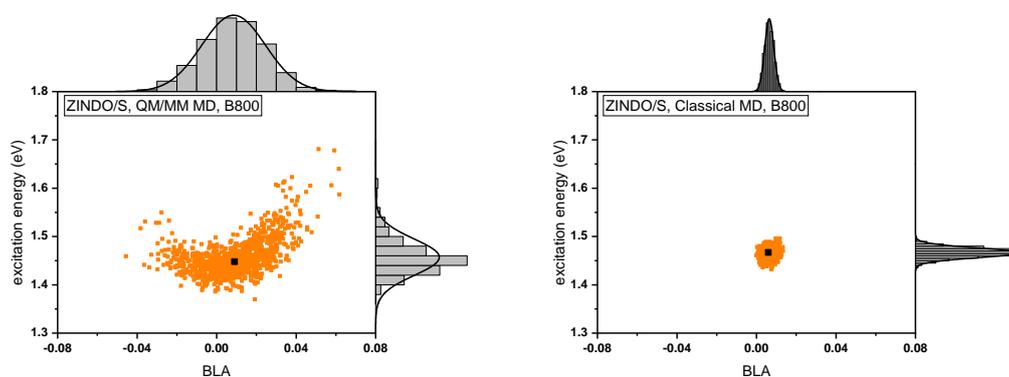
While the MM MD simulations obtain BLA values, which fluctuate around  $0.0\text{--}0.01 \text{ \AA}$ , the QM/MM obtained BLA values fluctuating around  $\pm 0.04 \text{ \AA}$ . This however, is a too large range for the BLA values of BChl a chromophores and indicate a wrong description of the ground state geometry by DFTB3. This is additionally reflected in Fig. 5.3 and 5.4, where the excitation energies of the BChl a chromophores in vacuum are computed. Due to the large range of the BLA values obtained by QM/MM a too broad absorption spectrum is observed. This however is in conflict with the experimental results, showing in the case of the LH2 complex spectra with a width of about  $0.1\text{--}0.14 \text{ eV}$ .<sup>66</sup> The width observed when using the

QM/MM sampled structures result in a range of 0.2 eV in the case of LC-DFTB.

The comparison of Fig. 5.3 to 5.4 reflects the different behavior of LC-DFTB and ZINDO/S. Both methods react differently on the various shape and BLA of the sampled BChl a chromophores. Further discussions are given in the next section.



**Figure 5.3** BLA (in Å) plotted vs. excitation energies (eV) of one BChl a chromophore of the B800 ring in vacuum using LC-DFTB. Left: sampled by QM/MM MD simulation; Right: sampled by MM MD simulation.



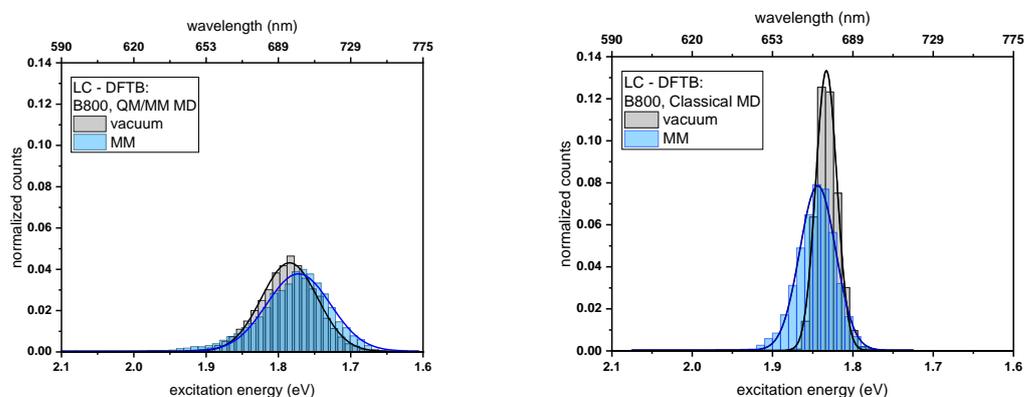
**Figure 5.4** BLA (in Å) plotted vs. excitation energies (eV) of one BChl a chromophore of the B800 ring in vacuum using ZINDO/S. Left: sampled by QM/MM MD simulation; Right: sampled by MM MD simulation.

### Environmental Effects

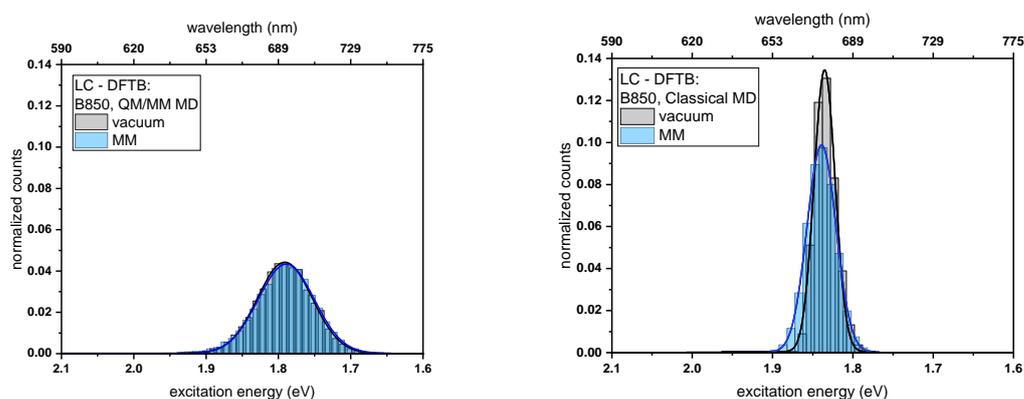
Analogous to the QM/MM optimized model, the effect of the protein environment on the excitation energies is shown here for the sampled structures by MM and QM/MM MD simulations.

The excitation energies are computed for 1000 snapshots using LC-DFTB and ZINDO/S for both B800 and B850 ring, respectively. All excitation energies are weighted by the oscillator

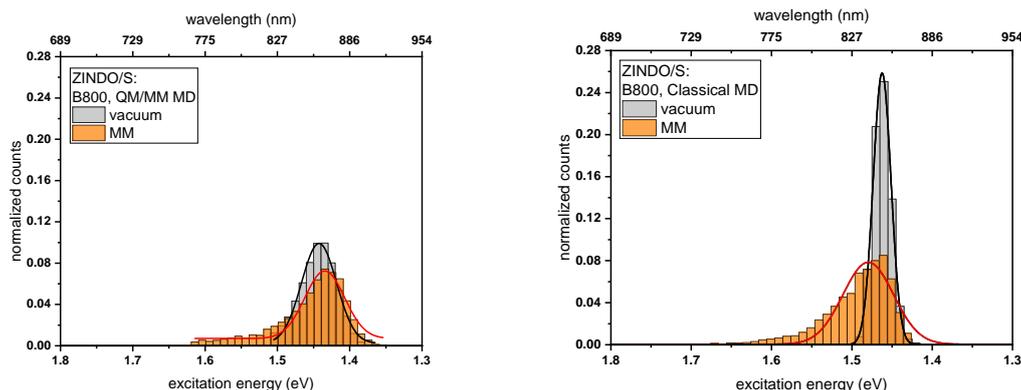
strength and plotted as histograms, see Fig. 5.5 and 5.6 for LC-DFTB and Fig. 5.7 and 5.8 for ZINDO/S. The absorption maxima are obtained with a fit using a Gaussian function and displayed in Tab. 5.5.



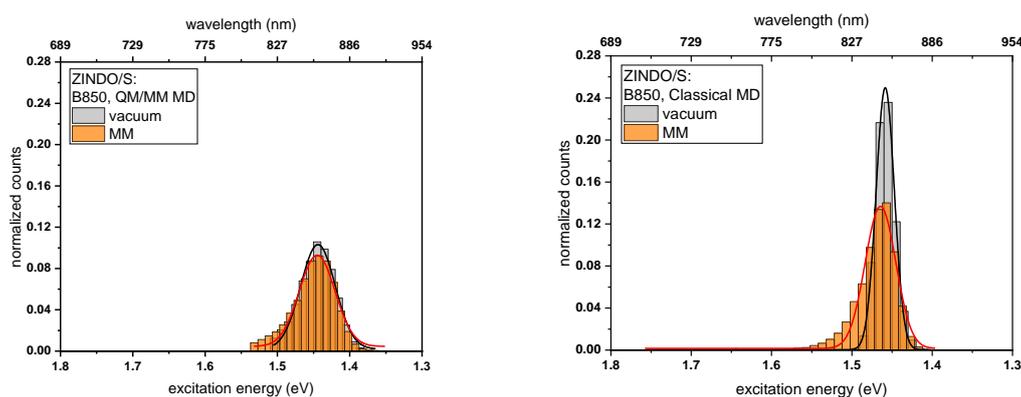
**Figure 5.5** Simulated absorption spectrum of the LH2 complex of the B800 ring. LC-DFTB is used for the computation of the excitation energies. The histograms are based on snapshot geometries of Left: QM/MM MD trajectory (1 ns length) and Right: MM MD trajectory (1 ns length). Plotted are the excitation energies weighted by the oscillator strength for (i) only the BChl a chromophore (vacuum) and (ii) with additional fixed MM point charges to account for the protein environment (MM). Gaussian functions are used to determine the corresponding maxima.



**Figure 5.6** Simulated absorption spectrum of the LH2 complex of the B850 ring. LC-DFTB is used for the computation of the excitation energies. The histograms are based on snapshot geometries of Left: QM/MM MD trajectory (1 ns length) and Right: MM MD trajectory (1 ns length). Plotted are the excitation energies weighted by the oscillator strength for (i) only the BChl a chromophore (vacuum) and (ii) with additional fixed MM point charges to account for the protein environment (MM). Gaussian functions are used to determine the corresponding maxima.



**Figure 5.7** Simulated absorption spectrum of the LH2 complex of the B800 ring. ZINDO/S is used for the computation of the excitation energies. The histograms are based on snapshot geometries of Left: QM/MM MD trajectory (1 ns length) and Right: MM MD trajectory (1 ns length). Plotted are the excitation energies weighted by the oscillator strength for (i) only the BChla chromophore (vacuum) and (ii) with additional fixed MM point charges to account for the protein environment (MM). Gaussian functions are used to determine the corresponding maxima.



**Figure 5.8** Simulated absorption spectrum of the LH2 complex of the B850 ring. ZINDO/S is used for the computation of the excitation energies. The histograms are based on snapshot geometries of Left: QM/MM MD trajectory (1 ns length) and Right: MM MD trajectory (1 ns length). Plotted are the excitation energies weighted by the oscillator strength for (i) only the BChla chromophore (vacuum) and (ii) with additional fixed MM point charges to account for the protein environment (MM). Gaussian functions are used to determine the corresponding maxima.

Analogously to the QM/MM optimized model no significant shift is obtained including the protein environment for both B800 and B850 ring by both methods. Tab.5.5 shows for both methods a red shift of the absorption maxima obtained by QM/MM compared to the absorption maximum obtained by MM. This in turn reflects, that the QM/MM obtained BChla geometries does not only have a larger BLA but also a different shape, since only a

larger BLA would lead to a blue shift as shown in Fig.4.8 of chapter 4.

In the case of LC-DFTB, the difference in the BLA due to the MM or QM/MM MD sampling is also reflected in the broadness of the histograms. The histograms of the MM MD trajectories show only half of the width than the histograms of the QM/MM MD trajectories. As already mentioned above, the results obtained by the QM/MM MD sampling is not in agreement to the experimental results, which show for B800 a width of about 0.1 eV and for the B850 ring a width of about 0.14 eV.<sup>66</sup>

In the case of ZINDO/S a complete different picture is obtained. The histograms observed by the MM MD simulations for the vacuum excitation energies are similar to those of LC-DFTB. However, the inclusion of the protein environment lead to a broadening of the histograms, which is not Gaussian distributed anymore, especially in the case of the B800 ring. This effect was also reported by previous studies.<sup>172</sup> Additionally, this reflects the overestimation of ZINDO/S when including the protein environment, as discussed in chapter 4. The results obtained by ZINDO/S are thus also not in agreement with the experimental observed results.

Furthermore, the effect of the QM/MM sampling with respect to the higher BLA of the BChl a geometries is not obtained. An only slightly broadening compared to the histograms of the MM sampled structures is obtained. This indicates the importance of the investigation of the BLA effect by both methods. Previous studies performing QM/MM MD simulations using PM6 and ZINDO/S on a water-soluble chlorophyll-binding protein (WSCP) reported an improvement by the QM/MM description.<sup>184</sup> The results of this chapter show however, that the combination of QM/MM sampling with ZINDO/S for subsequent excitation energy calculations should be considered with caution.

**Table 5.5** Excitation energies (in eV) of BChl a monomers of the LH2 complex in vacuum and with protein environment using LC-DFTB and ZINDO/S along the MM and QM/MM MD trajectory. The maximum is obtained by a Gaussian fit of the excitation energies for the B800 and B850 ring, respectively.

		LC-DFTB			ZINDO/S		
		vacuum	protein	shift	vacuum	protein	shift
QM/MM	B800	1.78	1.79	-0.01	1.44	1.43	+0.01
	B850	1.79	1.79	$\pm 0.00$	1.44	1.44	$\pm 0.00$
MM	B800	1.83	1.84	+0.01	1.46	1.48	+0.02
	B850	1.84	1.84	$\pm 0.00$	1.46	1.47	+0.01

### Exciton Coupling

The high amount of BChl a chromophores in the LH2 complex needs a multichromophoric QM region if one want to calculate the Coulomb couplings over the trajectory. However, the here performed QM/MM MD simulations contain only a single BChl a chromophore in the QM region, respectively. Therefore, the sampled structures by the QM/MM MD simulations

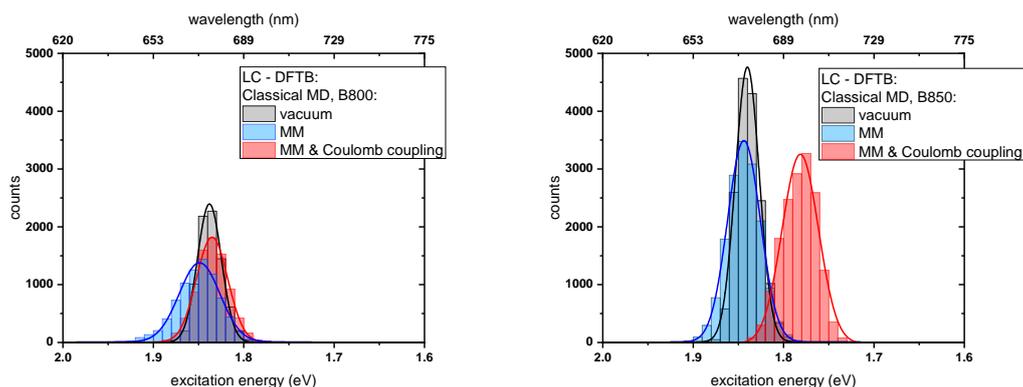
are excluded from the further computation of Coulomb couplings.

For the computation of the Coulomb couplings over the MM MD simulations, the procedure analogous to the QM/MM optimized model is performed. The averaged eigenvalues obtained of both rings are shown in Tab. 5.6. The respective Coulomb coupling values are in the same range as discussed for the QM/MM optimized model. A higher absorption maxima shift compared to the QM/MM optimized model is obtained when considering the couplings of all excitonic couplings of one ring, respectively. This results in a shift of about 86% of the experimental value.

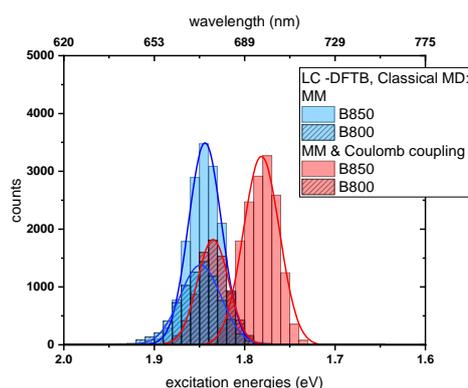
**Table 5.6** Lowest BChl a dimer energy in eV of LH2 resulted from diagonalization of the dimer exciton hamiltonian over the MM MD trajectory with the highest oscillator strength. The energies are averaged over the B800 and B850 ring, respectively. The maximum of a Gaussian fit of the eigenvalues for each BChl a chromophore are taken. Additional lowest energy, resulted by diagonalization of the (8x8) and (16x16) matrix for B800 and B850, without considering the oscillator strength. All excitation energies and Coulomb couplings include the protein environment and are calculated using LC-DFTB.

		B800	B850	B850 shift
LC-DFTB	monomer	1.84	1.84	-0.00
	dimer	1.84	1.78	-0.06
	complete ring	1.82	1.74	-0.08
Exp.		1.55	1.46	-0.09

To visualize the effect of the Coulomb couplings on the absorption maxima shift, histograms for both rings are plotted in Fig. 5.9 and 5.10. In these histograms the Coulomb couplings of the dimers are considered. In the case of the B800 ring, the inclusion of the protein environment leads to a slightly blue shift, while including additionally the Coulomb couplings a slightly red shift is obtained. The effect of the Coulomb couplings is significant visible in the case of the B850 ring showing a red shift of 0.06 eV.



**Figure 5.9** Simulated absorption spectrum of the LH2 complex of the B800 ring (left) and the B850 ring (right). LC-DFTB is used for the computation of the excitation energies and Coulomb couplings. The histograms are based on snapshot geometries of MM MD trajectory (1 ns length). Plotted are the excitation energies for (i) only the BChl a chromophore (vacuum) and (ii) with additional fixed MM point charges to account for the protein environment (MM) and (iii) including Coulomb couplings. Gaussian functions are used to determine the corresponding maxima.



**Figure 5.10** Simulated absorption spectrum of the LH2 complex, the histograms including the protein environment and Coulomb couplings of both rings are replotted in one diagram to visualize the resulted shift of B850 to B800.

Even though the exciton coupling benchmark in the previous chapter report an overestimation of LC-DFTB Coulomb couplings for the artificial system, the obtained shift by LC-DFTB is slightly underestimated with respect to the experiment. However, it has to be mentioned, that the application of the biological system obtained not an overall overestimation of the Coulomb couplings. Quantitatively, the values are sometimes over, under or equal compared to the supermolecule calculations using CAM-B3LYP and LC-DFTB. So that some error cancellation might be observed here. However, several other effects are neglected here. Since this is a primary test of the performance of LC-DFTB on such proteins not all influences are

taken into account, which are: the exciton coupling between both rings, the inclusion of the carotenoids and the possible polarization effects by the protein environment.

Furthermore, recent studies suggested the occurrence of charge transfer effects when investigating the LH3 complex<sup>42,169</sup> The absorption spectrum of the LH3 complex shows two maxima at 800 nm and 820 nm, while the latter corresponds to the B550 ring of the LH2 complex. The only difference in both LH complexes is obtained by the number of the chromophores. However, the same biological species from which the LH3 complex is obtained, contain also a LH2 complex with the same absorption maxima then the one used in this work. So that this absorption maxima shift was suggested to result by different hydrogen bond networks and charge-transfer effects.<sup>169</sup> A recent study on the absorption spectrum of the LH3 complex<sup>45</sup> using also LC-DFTB obtained an overestimation compared to the experimentally observed shift. However, the shift is similar to the one obtained in this work. Hence, the extension of the Hamiltonian considering also charge-transfer effects might give an improved insight of these behaviors.

## 5.5 Conclusion

In this chapter, the benchmark study of LC-DFTB of chapter 4 is extended to its performance on rhodopsins and LH complexes. Especially, when investigating such biological systems sampling is also required to account e.g. for the chromophore and protein environment fluctuation.

The application of LC-DFTB as well as LC-DFT functionals on the QM/MM optimized rhodopsin models reflects the same weakness as obtained by the benchmark study on retinal as they cannot correctly describe the size of the shift of absorption maxima between proteins. However, the LC-DFT functionals improve over standard DFT functionals, such as GGA or hybrid, as they could provide an improved description of charge transfer effects within the protein produced by the counterions and the hydrogen bonding network. The sampling of bR and ppR by QM/MM MD simulation with subsequent excitation energy computation using LC-DFTB and OM2/MRCI reveals the same results as the QM/MM optimized models. Since these chosen rhodopsins have a more strong hydrogen bond network within the active site only small changes are obtained through sampling. LC-DFTB gives a too small width of the sampled absorption spectrum compared to OM2/MRCI. Hence, OM2/MRCI is still the method of choice for the investigation of color-tuning in rhodopsins.

In the case of the LH2 complex, LC-DFTB reflects also the same properties for the description of the excited state properties of BChl a as obtained by the benchmark study. The QM/MM MD simulation on the LH2 complex sampled BChl a geometries, which differ significantly in the BLA compared to the MM MD sampled BChl a geometries. This is also reflected in the excitation energies of the BChl a chromophores showing a too broad absorption spectrum compared to the experimental one. However, this is only obtained by LC-DFTB,

since ZINDO/S shows no significant difference between QM/MM and MM MD sampled absorption spectrum, which might occur from error cancellation. Further studies are necessary to figure out, why QM/MM MD simulations lead to a complete different description of the BChl a geometries and whether this is only an error by DFTB3 or whether in this case well parameterized forcefields obtain better results. This is important, since especially QM/MM MD simulations are desirable in the research field of LH complexes.

The protein environment does not effect much on the absorption maxima shifts of the LH2 complex. However, the inclusion of the Coulomb couplings using LC-DFTB obtained a shift of about 86% of the experimental value. At this point an overall conclusion about reasons for the shifts can not be made, since several effects like polarization, or the Coulomb couplings between both rings as well as the influences of the carotenoids are not considered.

In general, LC-DFTB shows here its advantage to compute thousands of sampled structures due to its low computational effort and is thus promising for the study of LH complexes. Besides the points already mentioned in the conclusion of chapter 4, further benchmarks and improvement in the description of the LC-DFTB Coulomb couplings are necessary, since it is high desirable to have a method being independent on the experiment in order to get information about the physical influences on the Coulomb couplings, to yield an improved understanding of color-tuning effects in LH complexes. Furthermore, this study has to be also extended in order to consider additional effects, which might also play an important role, like polarization effects or possible charge-transfer effects as suggested in recent computational studies investigating the LH3 complex.<sup>42,169</sup>

---

## The First Intermediate State of Channelrhodopsin-2 Wild-Type

### 6.1 Introduction

Channelrhodopsin-2 wild-type (ChR2-WT) as introduced in chapter 2 was the first light-gated ion channel found in the green algae *Chlamydomonas reinhardtii*<sup>155,156</sup> and permeable for cations ( $\text{Na}^+$ ,  $\text{K}^+$ ,  $\text{Ca}^{2+}$ ). It exhibits a dimeric structure embedded in a membrane, where each monomer consists of seven transmembrane  $\alpha$ -helices (A-G) forming an internal pocket. The retinal chromophore is covalently bound to a lysine residue of the seventh helix (G) through a protonated Schiff base ( $\text{RSBH}^+$ ).

As common feature to other microbial rhodopsins, a photocycle is triggered after light activation of ChR2, see Fig. 2.10 in chapter 2. During the photocycle of ChR2 several reaction processes proceed like hydrogen bond changes, reorientation of amino acids followed by rearrangement of helical backbones and proton transfer reactions. All these processes are necessary for the channel functionality as opening, conductivity, closing and the back formation to the dark state.

Due to the lack of a crystal structure, computational studies were based on the hybrid C1C2 X-ray structure<sup>100</sup> or on homology models.<sup>13,120,224</sup> The hybrid-type C1C2 is a chimera of ChR1 and ChR2 and comprised of the first 5 helices (A-E) of ChR1 and the last 2 helices (F-G) of ChR2. The homology model as used in this work, is based on the C1C2 X-ray structure and obtained by sampling with classical force field (MM) MD simulations and structural refinement with combined QM/MM MD simulations.<sup>224</sup> Both, the C1C2 structure and the homology model<sup>224</sup> agree in common hydrogen bond patterns, orientation of charged residues and the active site. The main difference is the water density, since the homology model shows a higher water density in the protein core than the C1C2 X-ray structure.<sup>225</sup> Even though, ChR1 and ChR2 show a sequence identity of about 65%, recently electrophysiological measurements and FTIR spectroscopy obtained differences in the mechanism and function of

C1C2 compared to ChR2.<sup>87</sup>

In 2017, the X-ray structure of ChR2 was resolved<sup>217</sup> allowing now subsequent computational analysis. Thus, the homology model can be evaluated for its ability to investigate such proteins. The comparison between the X-ray structure of ChR2-WT and the C1C2 structure revealed the main differences in helices length and orientation of the helices. Especially helix A is about 7 Å longer in ChR2-WT than in C1C2. This might cause different hydrogen bond networks leading to different functionality of the channels. Furthermore, the water density is higher in the ChR2-WT X-ray structure than in the C1C2 chimera.<sup>217</sup>

The homology model<sup>225</sup> already showed a higher water density than in the C1C2 chimera, but even higher than in the ChR2-WT X-ray structure.<sup>232</sup> The main water distribution in the homology model is observed at the cytoplasmic side preserved by a cluster of hydrophilic residues (E82, E83, H134, H265, R268) as well as along helix B containing a large number of charged amino acids, while residues S63 and E90 separated it into two parts.<sup>224,225</sup> An extensive sampling of the ChR2-WT X-ray structure of up to 600 ns using MM MD simulations, showed an increase water density in the protein core as present in the homology model.<sup>232</sup> The reason for this observation lies in the harsh experimental conditions to which the protein is exposed when the crystal structure is determined. This can lead to changes in the fine structure of the protein, which was already shown in previous studies on bR,<sup>233</sup> or to a lower water density within the protein leading to a more rigid protein structure. However, the role of water was already highlighted as important for the functionality of rhodopsins.<sup>67,88</sup>

### 6.1.1 The Dark State

The dark state of the ChR2-WT represents the state without light contact. Structural changes differ often in only few kcal/mol making an initial well characterized dark state structure necessary for further investigation of the photocycle intermediates. Therefore, an extensive investigation of several structural motifs of ChR2-WT in the dark state was performed based on a MM as well as a QM/MM approach on both the homology model and the X-ray structure.<sup>73,232</sup>

The comparison between the homology model and the X-ray structure showed similarities in the binding pocket of retinal with respect to amino acid orientations and the hydrogen bond network, see Fig. A.3 in the appendix. The main differences are found in the water density in the active site and the orientation of the residue E90. In the following, the properties of the structural motifs of the dark state are briefly summarized.

#### The Active Site

The absorption spectrum of ChR2-WT in the dark state exhibits a multi-peaked spectrum<sup>32</sup> in contrast to the one-peak spectrum of bR. The active site of bR is characterized by the typical pentagonal hydrogen bond network.<sup>198,234</sup> In general, the absorption spectra of rhodopsins as

discussed in chapter 4 and 5 depend on the retinal geometry, the amino acids in the binding pocket as well as the hydrogen bond network. QM/MM MD sampling of the homology model combined with subsequent computation of excitation energies using SORCI and OM2/MRCI revealed a flexible active site structure.<sup>73</sup> This is reflected in different hydrogen bonding pattern of the RSBH<sup>+</sup>. The RSBH<sup>+</sup> forms either a hydrogen bond to the counterions E123 and D253 or to a water molecule, while the bonds to E123 or to the water molecule are preferred.<sup>73</sup> The flexible active site mainly accounts for the multi-peak absorption fine spectrum. The X-ray structure confirms the flexible hydrogen bonding pattern of the RSBH<sup>+</sup>, when the water density is higher in the active site than experimentally determined.<sup>232</sup> The experimental obtained X-ray structure suggested, that the RSBH<sup>+</sup> forms hydrogen bonds only to the two counterions E123 and D253.<sup>217</sup> However, QM/MM MD simulations showed, that the RSBH<sup>+</sup> forms about 80 % no hydrogen bond, when only the experimental obtained water molecules are present in the active site. This would not correspond to the multi-peak absorption fine spectrum as well as to Raman spectroscopy results predicting a strongly hydrogen bonded RSBH<sup>+</sup> to a water molecule.<sup>88,152,232</sup> Additionally, the temperature effect was also investigated performing MM MD simulations at 100 K on the homology model and subsequent comparison to experimental NMR results.<sup>20</sup> Due to the decreased kinetic energy only rigid hydrogen bonding patterns of the active site were obtained as found by the NMR studies.<sup>73</sup>

Furthermore, two different E123 conformations were found: (i) an upward to the cytoplasmic side and (ii) a downward orientation to the extracellular side. The upward orientation is slightly higher stable with about 2.5 kcal/mol than the downward orientation as obtained by well-tempered metadynamic simulations using MM MD simulations.<sup>73</sup> This orientation pattern has also an indirect but important influence on the hydrogen bond pattern of the RSBH<sup>+</sup>, since the downward orientation of the E123 is not hydrogen bonded to the RSBH<sup>+</sup>. Additionally, the flip of E123 is also obtained by the X-ray structure with a higher water density.<sup>232</sup>

## The DC Gate

The DC gate plays a crucial role in the channel opening and closing process, since the mutant C128T shows a significant change of the channel kinetics. C128T exhibits an extended life time of the conducting state of about 10<sup>2</sup>- to 10<sup>5</sup>- fold.<sup>182,203</sup> The residues involved in the DC gate are C128 and D156, which are connected via a water molecule as previously resulted by QM/MM MD simulations and vibrational frequency analysis<sup>225,228</sup> and later confirmed by the Chr2 X-ray structure.<sup>217</sup> At this point, it has to be mentioned, that the C1C2 structure contains a direct bond between D156 and C128, which is also found in the C128T mutant.<sup>73,217</sup> The water in the DC gate however might play a significant role in the channel functionality. On the one hand, it stabilizes helices C and D and on the other hand it might be involved in

the reprotonation step of the protonated RSB, since D156 is suggested to be the proton donor of the protonated RSB, as resulted by FTIR measurements.<sup>134</sup> Furthermore, the hydrogen bond pattern of the active site is changed dependent whether a direct or a water-bridged DC-Gate is present. The C128T mutant showed a preferred hydrogen bond of the RSBH<sup>+</sup> to a water molecule, while ChR2-WT containing the water-bridged DC-Gate preferred two pattern, the RSBH<sup>+</sup> hydrogen bonded to E123 or a water molecule.<sup>232</sup>

### The Role of E90

In the X-ray structure, E90 is oriented downward towards the extracellular side and hydrogen bonded via a water molecule to K93, E123 and D253.<sup>217</sup> This is the main difference compared to the homology model, containing an upward orientation towards the cytoplasmic side of E90 bonded to N258. The energetically preferred downward orientation of E90 was also obtained in the homology model by well-tempered metadynamic simulations based on MM MD simulations showing a barrier of 6.5 kcal/mol between both conformations.<sup>232</sup> Excitation energy calculations on the X-ray structure revealed no large differences compared to the excitation energies obtained with the homology model containing an upward orientation of E90.<sup>232</sup>

Mutations of E90 result in changes of ion selectivity. Hence, E90 was set to be also relevant for the channel functionality.<sup>71</sup> The deprotonation state of E90 during the photocycle is controversial discussed in the literature. Based on MM MD simulations an early deprotonation of E90 was suggested as well as a downward flip of E90 directly connecting with the channel preparation for opening.<sup>120</sup> However, FTIR results revealed a late deprotonation of E90 in the P<sub>3</sub> intermediate state.<sup>181</sup>

### 6.1.2 Isomerization of all-*trans* Retinal

The isomerization of all-*trans* retinal to 13-*cis* retinal upon photon absorption is the primary event in the ChR2 photocycle. Recently, a CASPT2//CASSCF/MM study on the retinal isomerization pathway was performed considering different initial dark state structures, i.e. different RSBH<sup>+</sup> hydrogen bonding pattern, as found in the homology model of ChR2-WT.<sup>74</sup> Among the three initial dark state structures as defined before, the RSBH<sup>+</sup> hydrogen bonded to E123 pattern was found to give the most probably isomerization pathway, due to the low energy barrier on the excited state. Subsequent vibrational frequency calculations of the obtained 13-*cis* retinal coincided with the experimental reported vibrational bands.<sup>74</sup>

### 6.1.3 Experimental Characterization of the First Intermediate State

The photocycle of ChR2-WT exhibits several intermediate states, which are characterized by spectroscopic techniques based on UV/Vis, FTIR or Raman spectroscopy.<sup>50, 116, 133–135, 216</sup>

They give an insight about the retinal configuration, the protonation state of retinal and other amino acids, structural rearrangements or changes in water distribution.

The first intermediate state  $P_1$  is involved in the pre-gating process and prepares the channel for the deprotonation of the  $RSBH^+$  and subsequent opening of the channel.  $P_1$  is formed within 3 ps after retinal isomerization.<sup>133–135</sup> The absorption maximum is about 30 nm red-shifted compared to the dark state as obtained by UV/Vis spectroscopy.<sup>134</sup>

One main difference compared to bR or other rhodopsins is found in the amide I region showing an intense negative band at  $\sim 1664\text{ cm}^{-1}$ , which is much larger than in other rhodopsins and is believed to result from an ultrafast deformation of a transmembrane helix due to a steric clash of the isomerized retinal with its surrounded environment.<sup>134,135,161,183</sup> This large structural changes might be necessary for the formation of a larger pore, since in ChR2 cations are conducted while for example in bR only protons are pumped.<sup>183</sup> Furthermore, the formation or enlargement of water-filled pores inside the protein might be a prerequisite for cation permeation.<sup>132,134</sup>

Lórenz-Fonfría *et al.*<sup>135</sup> proposed a ChR2-WT photocycle with the first intermediate state divided into two substates  $P_{1a}$  and  $P_{1b}$  differing in the peptide backbone alteration and in the interhelical hydrogen bond between D156 and C128. FTIR spectroscopy could give further insights, since both substates cannot be distinguished by UV/Vis spectroscopy. The carboxylic region showed a blue shift of the  $1737\text{ cm}^{-1}$  band of the dark state to  $1745\text{ cm}^{-1}$  in  $P_1$ , which was assigned to the weakening of the H-bond of D156. Additionally, this reflects a difference compared to the other rhodopsins, since e.g. in bR a red-shift of this band with smaller intensity is observed for the first intermediate state ( $1740(-)/1733(+)\text{ cm}^{-1}$ ).<sup>28,133,153,180</sup> Furthermore, in ChR2-WT a broad band at  $1760\text{ cm}^{-1}$  is found in  $P_1$  originating also from the C=O stretch of D156. This band was assigned to the late  $P_{1b}$  substate obtained by time-resolved tunable quantum cascade laser (QCL) IR measurements<sup>135</sup> and suggested a transient weakening of the interhelical H-bond between D156 and C128. This might be caused by a rotation or movement of helices D and C at around 200 ns after the formation of  $P_{1a}$ .

Structural changes are also expected in the central gate due to a blue shift of the  $1717\text{ cm}^{-1}$  band to  $1728\text{ cm}^{-1}$ , which was assigned to hydrogen bond changes of the carboxylic group of E90.<sup>133,134</sup>

The suggested hydrogen bond changes in both, the DC gate and central gate, could be accompanied by alterations in the inter-helical contacts between helices A-B-G and C-D and a possibly change of the relative geometry of these helices.

The experimental results provide an insight into the structural changes that can occur in the first intermediate state. The cause of the obtained shifts however, cannot be figured out experimentally. There exist already computational studies, but these are based on the C1C2 structure using MM MD simulations.<sup>13,37,208</sup> Additionally, these studies use rather harsh conditions to model the isomerized retinal, neglecting effects of energy barriers of the

retinal excited state isomerization pathway. This can have drastic consequences on the protein fine structure of the  $P_1$  state as well as on the pre-gating process. As for the dark state, the investigation of the first intermediate state requires different methods based on both a MM and a QM/MM approach. Using these methods, one can account also for spectroscopic silent intermediates, multiple conformations or other structural changes, which might be important for the next steps, like the deprotonation of the  $RSBH^+$  and the preparation for channel opening.

Hence, in this chapter the investigation of the first intermediate state  $P_1$  of ChR2-WT is presented. QM/MM as well as MM MD simulations are performed to sample several substates of  $P_1$ , which are further used for the computation of UV/Vis and vibration spectra. This makes a direct comparison with experimentally observed results possible. Both the X-ray structure and the homology model are used to investigate  $P_1$ . This allows a direct comparison of structural differences as existing in the central gate, and their influences on the pre-gating process.

## 6.2 Computational Details

### 6.2.1 Preparation of the $P_1$ State

#### The Early Substate $P_{1a}$

Both, the homology model<sup>224</sup> based on the X-ray structure of the C1C2-hybrid (PDB code: 3UG9)<sup>100</sup> and the X-ray structure of ChR2-WT (PDB code: 6EID)<sup>217</sup> were used for the study of the first intermediate state  $P_1$ . Both protein structures are constructed as a dimer embedded in a POPC (1-Palmitoyl-2-oleoylphosphatidylcholine)-bilayer as lipid membrane and surrounded by water molecules using the TIP3P<sup>95</sup> water model. Both model setups are provided by Ref. 73,232 and references therein.

In the case of the homology model, two structures of the dark state<sup>73</sup> were chosen: (i) a less equilibrated structure with a higher water density and (ii) a longer equilibrated structure with a lower water density. The first motif (i) is hereinafter referred to as system HW and the second motif (ii) as system LW. Both systems show the same hydrogen bond (H-bond) pattern of the active site, i.e. the  $RSBH^+$  is hydrogen bonded to E123. The use of this motif is consistent with the study of the retinal isomerization pathway,<sup>74</sup> which shows that the preferred path proceeds from this hydrogen bonding pattern.

The investigation of the dark state using the X-ray structure revealed a higher water amount near the  $RSBH^+$  leading to more reliable results.<sup>232</sup> Therefore, the X-ray structure experimentally obtained with less water in the active site was excluded here. For the investigation of  $P_1$  different structural motifs were chosen, as listed in Tab. 6.1, using the sampled X-ray structure containing a higher water amount.

The CASPT2//CASSCF/MM study on the retinal isomerization pathway on the homology

model of ChR2-WT by Guo *et al.*<sup>74</sup> received a 13-*cis* retinal conformation with a C<sub>12</sub>-C<sub>13</sub>=C<sub>14</sub>-C<sub>15</sub> dihedral angle twisted about 20°. This was also suggested by a HOOP (Hydrogen Out Of Plane) mode at 988cm<sup>-1</sup> by FTIR results similar to those observed by bR.<sup>134</sup> This 13-*cis* retinal conformation was taken and aligned to all dark state structures of both, the homology model and the X-ray structure. The alignment was performed with the *measure fit* command provided by VMD.<sup>85</sup>

**Table 6.1** Several structural motifs chosen for the investigation of P<sub>1a</sub> originating from QM/MM sampled X-ray structures.<sup>232</sup>

structural motif	active site	
	E123 orientation	RSBH <sup>+</sup> H-bond
E123 <sup>down</sup> -RSBH <sup>+</sup> ... -O(D253)	down	D253
E123 <sup>down</sup> -RSBH <sup>+</sup> ... H <sub>2</sub> O	down	H <sub>2</sub> O
E123 <sup>up</sup> -RSBH <sup>+</sup> ... H <sub>2</sub> O	up	H <sub>2</sub> O
E123 <sup>up</sup> -RSBH <sup>+</sup> ... -O(E123)	up	E123

### The Late Substate P<sub>1b</sub>

Several MM MD simulations were performed, details see section 6.2.3, to sample several possible P<sub>1</sub> substates. Three characteristic motifs were chosen to perform subsequent QM/MM MD simulations. All structures originated from the X-ray structure. The three motifs are characterized by different DC gate motifs, see Tab. 6.2.

**Table 6.2** Several structural motifs chosen for the investigation of P<sub>1b</sub> originating from MM sampled X-ray structures.

structural motif	DC-Gate
D156...H <sub>2</sub> O...C128	two H <sub>2</sub> O bonded to D156
(D156)O...S(C128)	direct DC-Gate (D156 bonded to C128)
(D156)O...O(C128)	direct DC-Gate (D156 bonded to C128 backbone)

### 6.2.2 QM/MM MD Simulation

Previous studies on bR showed the importance to describe the active site containing the retinal chromophore and the counterion as well as the nearby water molecules with QM methods, here DFTB.<sup>49,225,233</sup> In general, a well chosen QM region is important for retinal proteins to consider effects, like charge transfer, protein polarization and dispersion interaction.<sup>220</sup> For the investigation of the P<sub>1</sub> state, the QM region has to be adjusted, since the RSBH<sup>+</sup> of the isomerized retinal changed its orientation, pointing now towards the cytoplasmic side. In the case of the homology model, several QM regions were chosen, see Tab. 6.3. The DC gate and

the central gate were also included in the QM region due to their potential importance in the  $P_1$  state as obtained by experimental results, see Tab. 6.3.

**Table 6.3** Several QM regions chosen for the study of the early  $P_{1a}$  state.

abbreviation	QM region
QM	retinal, K93, E123, D253 and H <sub>2</sub> O
QM-TS	retinal, K93, E123, D253, T127, S221 and H <sub>2</sub> O
QM-DC <sup>1</sup>	retinal, K93, E123, D253, T127, S221, C128, D156 and H <sub>2</sub> O
QM-E90 <sup>1</sup>	retinal, K93, E123, D253, T127, S221, C128, D156, N258, E90 and H <sub>2</sub> O

<sup>1</sup> this QM region is only used for system LW.

The QM/MM study of the dark state based on the X-ray structure revealed only a stable E90 downward conformation, when using a large QM region, i.e. the inclusion of E90 into the QM region. Hence, the same QM region was chosen for the investigation of the  $P_1$  state in the case of the X-ray structure. This QM region is similar to the one of the homology model (QM-E90), see Tab. 6.3 but without the residue N258.

The QM/MM bond was set between the  $C_\alpha$  and  $C_\beta$  of each amino acid side chain with a standard link atom approach. Restraints were applied on the oxygen atoms of the QM water molecules to avoid an interchange between QM and MM water molecules, leading otherwise to convergence problems due to the unbalanced description. For the QM part in the QM/MM MD simulations, the DFTB3/3ob method (extended self-consistent-charge Density-Functional Tight-Binding)<sup>63-65</sup> was used, implemented in the GROMACS program package.<sup>118</sup> DFTB3/3ob is a reliable model for the description of hydrogen-bound structures as they occur in rhodopsins. It is able to describe hydrogen-bonded networks with similar accuracy as full DFT calculations performed with medium sized basis set.<sup>63</sup> For the MM part the CHARMM36 force field<sup>84</sup> was used. All QM/MM MD trajectories have a time step of 1 fs and a length of 1 ns, respectively. For each start structure of  $P_{1a}$  and  $P_{1b}$ , five QM/MM MD simulations were performed to account for different starting velocities and therefore a better statistical representation. Four QM/MM MD simulations of the X-ray structure failed due to steric clashes and were further on excluded from subsequent analysis.

All QM/MM MD simulations for every start structure were directly performed after the alignment of the 13-*cis* retinal conformation. In the case of system HW however, an equilibration was previously necessary.

The equilibration procedure was as follows: Firstly, a MM minimization was performed of the potential energy with the steepest descent algorithm (1000 steps) by keeping the retinal coordinates frozen. The frozen coordinates are necessary in order to keep the twisted 13-*cis* retinal conformation during the equilibration. Additionally, the heavy atoms of the protein and the lipid membrane were restrained to their initial positions by harmonic potentials with the force constant of  $1000 \text{ kJ mol}^{-1} \text{ nm}^{-1}$ . This is a usual procedure to keep the protein intact,

while equilibrating the environment. For the next step, a 250 ps long NVT equilibration was performed using the Berendsen thermostat<sup>21</sup> at 300 K. Finally, an overall (i.e. restraints on protein and lipid heavy atoms were removed) 250 ps long NPT equilibration was done using the Parinello-Rahman barostat<sup>173</sup> at a pressure of 1 bar.

### 6.2.3 MM MD Simulation

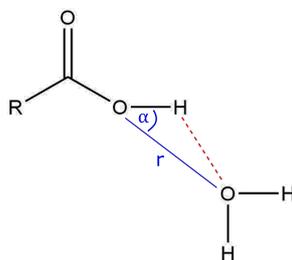
In order to sample several  $P_1$  substates and due to the fact that the lifetime of  $P_1$  is about 400 ns long, several MM MD simulations each with a length of 400 ns were performed. The same equilibration procedure as done for the system HW, but using the Nosé-Hoover thermostat<sup>83,168</sup> were performed. For system HW, system LW (homology model) and for the X-ray structure four MM MD simulations were performed, respectively. In order to additionally consider different sampling, the last frame of two QM/MM MD simulations was chosen in order to carry out MM MD simulations with these.

All MM MD simulations were carried out with the GROMACS package (version 2016.3)<sup>80</sup> and the CHARMM36 force field<sup>84</sup> with appropriate parameters for the lipid membrane.<sup>104</sup> During all simulations periodic boundary conditions were applied and the long-range interactions by means of the Particle-Mesh Ewald method were considered.

### 6.2.4 Analysis of the Trajectories

The analysis of the QM/MM MD trajectories were performed with the GROMACS package,<sup>118</sup> while the MM MD trajectories are analyzed with the GROMACS package (version 2016.3).<sup>80</sup> For the MM MD simulations the RMSD (root-mean-square deviation) for each helix is calculated.

The hydrogen bonds were computed under a geometrical criteria, which states that the distance between donor and acceptor is less or equal 0.35 nm and the angle between the bond of hydrogen-donor and the direction donor-acceptor is less or equal  $30^\circ$ . An example is shown in Fig. 6.1.



**Figure 6.1** Geometrical criteria for the computation of the H-bond, with the distance  $r \leq 0.35$  nm and  $\alpha \leq 30^\circ$ . R=rest of an amino acid, here the amino acid is the donor (OH) and the water molecule the acceptor (O).

### 6.2.5 Excitation Energies

All excitation energies were computed using the semi-empirical OM2/MRCI method as implemented in the MNDO2005 program package.<sup>113,226</sup> As discussed in chapter 4 and 5, the OM2/MRCI method has been extensively tested for its application to study color tuning effects in rhodopsins. It shows a systematic hypsochromic shift of 0.3 eV compared to SORCI.<sup>59,73,82</sup> This shift was obtained by using only the chromophore retinal in the QM region. Additionally, a large active space is possible due to the less computational costs. For the OM2/MRCI calculations an active space of (20,20) is used as previously proposed by Guo *et al.*,<sup>73</sup> which leads to an accurate description of the excitation energies. This means 10 occupied  $\pi$ -orbitals, 10 unoccupied  $\pi^*$ -orbitals and 20 electrons are considered in the calculation. Additionally, within the active space all single and double excitations are considered. The OM2/MRCI excitation energy calculations were performed for each snapshot obtained by the QM/MM MD trajectory, i.e. 1000 snapshots for each QM/MM MD trajectory of 1 ns length.

Furthermore, as discussed in chapter 5, protein polarization can be included by polarizable force fields to gain an improved description. However, this procedure is very costly and cannot be used for this study as it focuses on the sampling of different substates. The use of standard fixed charge QM/MM techniques lead in this case to a blue shift of the excitation energies about of 0.1-0.2 eV.<sup>219,221</sup>

### 6.2.6 Vibrational Frequencies

For the investigation of structural changes in the DC gate, vibrational frequencies were calculated by applying the fourier transform of time-correlation functions (FTTCF) method.<sup>98</sup> The procedure was as follows: Firstly, a geometry optimization with CHARMM was performed using the CHARMM36 force field for the MM region. For the QM region the DFTB3/3ob method was applied. The 3ob parameters were chosen. An extensive benchmark of different parameter sets and DFT functionals were performed by Welke *et al.*<sup>228</sup> revealed that the 3ob parameter set yields the smallest error with respect to PBE and can reproduce qualitatively the experimentally observed shifts.

The QM regions were differently chosen dependent which structure was analyzed. The QM region contained the DC gate, i.e. D156 and C128 and additional the water molecule in the case of a water-bridged DC gate. Structures were also found, where W223 forms a hydrogen bond to D156. In this case W223 was also added to the QM region. The QM/MM setup was performed as described in section 6.2.2.

For the geometry optimization the conjugated-gradient method was applied (1000 steps) until a gradient threshold of  $1 \times 10^{-3}$  a.u. was reached. Additionally, the Newton-Raphson (ABNR) minimizer was used, for a more precise minimization with a gradient threshold of  $1 \times 10^{-3}$  a.u. Subsequently, 100 independent QM/MM systems were created. Each system was equilibrated to a temperature of 300 K. After equilibration of the systems, short QM/MM MD

simulations of 20 ps lengths were performed with a time step of 0.5 fs, respectively. During the simulations the dipole moments were obtained. Subsequently, the fourier transform of the dipole moment autocorrelation function was calculated. At last, a correction factor was applied to account for quantum effects of nuclear motions. Finally, a spectrum was obtained averaged over 100 separate spectra.

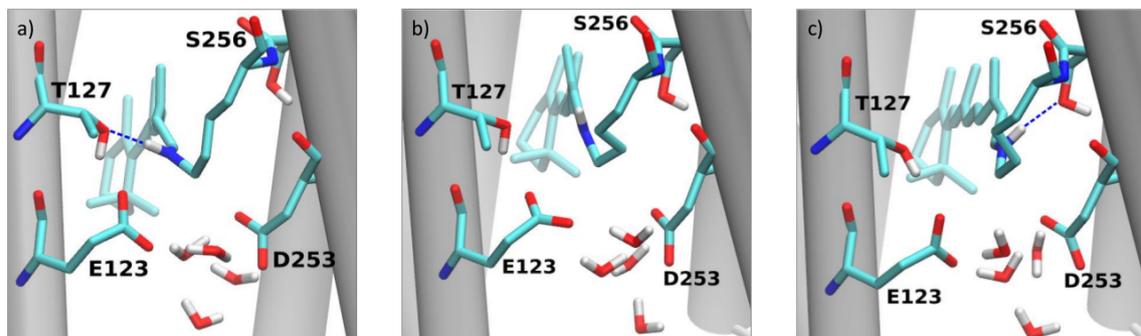
## Results and Discussion

### 6.3 QM/MM MD Simulations of the Early Substate P<sub>1a</sub>

The light activation of ChR2-WT triggers intermediate states, which represent parts of the photocycle. The primary event in the photocycle is the isomerization of the all-*trans* retinal upon photon absorption, forming a 13-*cis* retinal configuration. After 3 ps the first intermediate state P<sub>1</sub> is formed, more precisely the first substate P<sub>1a</sub> as suggested by FTIR experiments.<sup>135</sup> To investigate all structural changes triggered by the isomerized retinal QM/MM MD simulations were performed and the results are presented in the following sections.

#### 6.3.1 The Retinal Binding Pocket

##### Hydrogen Bonds of the RSBH<sup>+</sup>



**Figure 6.2** Three motifs of the RSBH<sup>+</sup> of the 13-*cis* retinal configuration found in all QM/MM MD simulations for both homology model and X-ray structure: a) H-bond: RSBH<sup>+</sup>  $\cdots$  O-(T127); b) no H-bond; c) H-bond: RSBH<sup>+</sup>  $\cdots$  O-(S256).

The retinal isomerization from all-*trans* to 13-*cis* configuration disrupts the H-bonds of the RSBH<sup>+</sup>, which was previously bound to either E123, D253 or a water molecule in the dark state. The QM/MM MD simulations of the early P<sub>1a</sub> state reveal still a hydrogen bonded RSBH<sup>+</sup> independent on the QM region or the start structure for both, the homology model

and X-ray structure. The RSBH<sup>+</sup> shows the following pattern: (i) either hydrogen bonded to residue T127 (RSBH<sup>+</sup> ... O-(T127)) or to (ii) S256 (RSBH<sup>+</sup> ... O-(S256)) or (iii) no H-bond, see Fig. 6.2. The percentage of the respective H-bond pattern are calculated and displayed in Tab. A.10 together with the dihedral angle (C<sub>12</sub>-C<sub>13</sub>=C<sub>14</sub>-C<sub>15</sub>) of retinal, which is discussed in the following section.

**Table 6.4** Percentage of the H-bond pattern (%) of RSBH<sup>+</sup> ... O-(T127) and RSBH<sup>+</sup> ... O-(S256) for the homology model and the X-ray structure. And the dihedral angle C<sub>12</sub>-C<sub>13</sub>=C<sub>14</sub>-C<sub>15</sub> (°) of the 13-*cis* retinal (in the table abbreviated as C<sub>13</sub>=C<sub>14</sub>). The values are averaged over all QM/MM MD trajectories, respectively.

Model	structural motif of the dark state	QM region	T127	S256	C <sub>13</sub> =C <sub>14</sub>
homology model	system HW <sup>a)</sup>	QM	0.0	43.0	-4.93
		QM-TS	3.0	8.7	+1.86
	system LW <sup>a)</sup>	QM	0.0	28.6	-1.89
		QM-TS	14.3	3.3	+14.0
		QM-DC	17.5	1.7	+16.4
X-ray	E123 <sup>down</sup> -RSBH <sup>+</sup> ... O(D253)	QM-E90	1.2	12.4	+3.85
			1.9	11.9	+4.39
			4.1	0.0	+12.7
			21.6	2.2	+17.2

<sup>a)</sup>E123<sup>up</sup>-RSBH<sup>+</sup> ... O(E123)

Differences are found in the percentage of the H-bond pattern. The investigation of different QM regions of the homology model reflects a main difference when both residues T127 and S256 located near the RSBH<sup>+</sup> are not included in the QM region. In this case no H-bond pattern of RSBH<sup>+</sup> ... O-(T127) is found. Interestingly, if the QM region is larger, i.e. includes also the DC gate or/and the central gate, the percentage of the H-bond pattern RSBH<sup>+</sup> ... O-(T127) increases over the H-bond pattern RSBH<sup>+</sup> ... O-(S256). Furthermore, the total percentage of the hydrogen bound RSBH<sup>+</sup> is reduced from the small QM region (30-40%) to the large QM region (~20%). This mainly reflects the importance of the chosen QM region as previously reported for bR and for the dark state study of Chr2-WT.<sup>73,225,233</sup>

The comparison of the homology model with the X-ray structure reveals similarities in the percentage of the H-bond pattern of the RSBH<sup>+</sup>, when the QM/MM MD trajectories starting with the same structural motif as presented in the dark state. Differences are obtained for the other structural motifs reflecting a higher amount of the RSBH<sup>+</sup> ... O-(S256) H-bond pattern or nearly no H-bond of the RSBH<sup>+</sup> as shown in Tab. A.10.

Experimental Raman spectroscopy results revealed a red-shift of the  $\nu$ C=N band in the

amid I region of about  $30\text{ cm}^{-1}$  from the dark state to the first intermediate state, while the magnitude of the shift is similar to bR.<sup>140,152</sup> This band was assigned to the  $\nu\text{C}=\text{N}$  stretching mode of the RSBH<sup>+</sup> and the shift indicates, that the H-bond of RSBH<sup>+</sup> as presented in the dark state is disrupted after retinal isomerization. A further FTIR spectroscopy study on the Chr2-E123T mutant, however suggested that the RSBH<sup>+</sup> is still hydrogen bonded even if weakly in the P<sub>1</sub> state due to the higher frequency of the  $\nu\text{C}=\text{N}$  band ( $1623\text{ cm}^{-1}$ ) in contrast to bR ( $1609\text{ cm}^{-1}$ ).<sup>135</sup> This might also be the case in the Chr2-WT, since the dark state bands of wild-type and mutant are about  $1655\text{ cm}^{-1}$ . However, the experimental observation cannot determine whether the RSBH<sup>+</sup> is still bound to a counterion despite the isomerization, or to another amino acid of the retinal binding pocket.

The QM/MM MD simulations of the P<sub>1a</sub> state found a still hydrogen bonded RSBH<sup>+</sup> and thus confirm the experimental assumption of a still weakly bound RSBH<sup>+</sup>. This is also in agreement with the percentage of the H-bond, since the large QM region reflects with  $\sim 20\%$  a not strongly hydrogen bonded RSBH<sup>+</sup>.

Cheng *et. al*<sup>37</sup> observed also a hydrogen bonded RSBH<sup>+</sup> in the P<sub>1</sub> state using MM MD simulations. However, they found only the H-bond pattern RSBH<sup>+</sup>  $\cdots$  O-(T127). Ardevol *et. al*<sup>13</sup> obtained a complete different H-bond pattern in the P<sub>1</sub> state, namely RSBH<sup>+</sup> hydrogen bonded to D253, using also MM MD simulation. These different observations mainly reflect the importance of the chosen isomerized retinal structure. While both previous computational studies neglect excited state effects on the retinal configuration, the retinal configuration chosen here, originates from a CASPT2//CASSCF/MM study on the retinal isomerization pathway,<sup>74</sup> taken thus different influences on the retinal configuration upon excitation into account. The results by Cheng *et. al*<sup>37</sup> are in accordance to our results, since the dihedral angle around the C<sub>13</sub>=C<sub>14</sub> of the 13-*cis* retinal used in their study is about  $25^\circ$  and thus similar to the dihedral angle of the retinal ( $20^\circ$ ) used in this work, see Tab. A.10. In contrast, Ardevol *et. al*<sup>13</sup> rotate the RSBH<sup>+</sup> during the isomerization procedure into the other direction, i.e. towards the counterion D253, analogous to the isomerization pathway observed in bR. However, Guo *et. al*<sup>74</sup> pointed out, that this isomerization pathway lead to a too high energy barriere, than observed for the other direction, i.e. towards the counterion E123. The detailed investigation of the RSBH<sup>+</sup> H-bond pattern is important, since it can have a major influence on further processes, as it will be discussed below, see section 6.4.2.

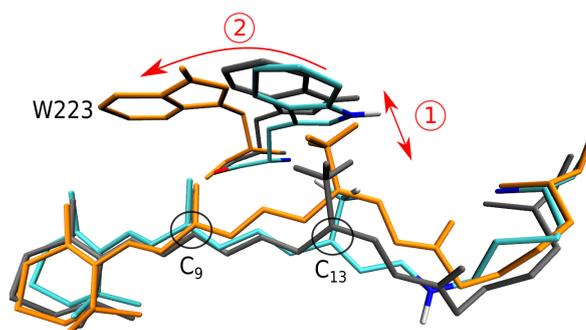
### Steric Clash of Surrounded Amino Acids with the Isomerized Retinal

The retinal isomerization study<sup>74</sup> observed a 13-*cis* retinal conformation with the RSBH<sup>+</sup> pointing towards T127 and a slightly twisted conformation with a C<sub>12</sub>-C<sub>13</sub>=C<sub>14</sub>-C<sub>15</sub> dihedral angle of about  $20^\circ$ . The study on the RSBH<sup>+</sup> H-bond pattern reflects, that the retinal is still slightly twisted when the RSBH<sup>+</sup> forms a H-bond. The dihedral angle C<sub>12</sub>-C<sub>13</sub>=C<sub>14</sub>-C<sub>15</sub> is measured and averaged over the QM/MM trajectories, and listed in Tab. A.10. As expected,

the obtained dihedral angle values behave analogously to the H-bond pattern of the RSBH<sup>+</sup> showing a more planar retinal, if the H-bond pattern RSBH<sup>+</sup> ··· O-(S256) is dominated.

The isomerized retinal can not only have influences on the RSBH<sup>+</sup> H-bond pattern as discussed above, but can lead also to structural changes in the binding pocket due to the steric interactions between the isomerized retinal and any surrounded amino acid. A reduced distance to any amino acid could subsequently lead to a steric clash. This was suggested by experimental studies, since FTIR experiments found an intense band in the amide I region (at  $\sim 1664\text{ cm}^{-1}$ ) much larger than obtained by other rhodopsins.<sup>134,135,161,183</sup> Because of the different intensity compared to other rhodopsins, they suggested a subsequent helix movement or deformation due to the steric clash.

Therefore, the structural changes in the binding pocket of the retinal are analyzed in all QM/MM MD simulations. In mostly all QM/MM simulations changes appear in the orientation of the indol ring belonging to the residue W223. In the dark state, the indol ring of W223 is located above the C<sub>13</sub>-methyl group of retinal. To have a more detailed insight, the distance between the center of mass (COM) of the C<sub>13</sub>-methyl group of retinal and the COM of the indol ring of W223 is measured in all QM/MM trajectory for both homology model and X-ray structure. The values are given in Tab. A.9 in the appendix. This distance varies about 3.8 – 4.0 Å in the dark state of the homology model. Dependent on the QM region, the homology model shows compared to the dark state a reduced distance of  $\sim 0.4 - 0.7$  Å for the small QM region (QM, QM-TS) for both system HW and LW or a slightly reduced distance of about  $\sim 0.2 - 0.4$  Å for the larger QM region (QM-DC, QM-E90) of system LW. In the larger QM region however a movement of the indol ring of W223 towards the C<sub>9</sub>-methyl group of retinal is found in three out of 10 QM/MM MD trajectories, see Fig. 6.3.



**Figure 6.3** Three retinal conformations are shown. Cyan: all-*trans* retinal in the dark state; Grey: 13-*cis* retinal at the beginning of the QM/MM simulation; Orange: 13-*cis* retinal, when the indol ring of W223 is moving towards the C<sub>9</sub>-methyl group of retinal; Step ①: reduced distance triggers step ②: movement of indol ring of W223.

In the X-ray structure, the same observation is made in five out of 16 QM/MM MD trajectories. The QM/MM MD trajectories without the movement of the indol ring yield a reduced averaged

distance of about 0.7–0.8 Å. The slightly reduced distance could trigger the movement of the indol ring of W223 from the C<sub>13</sub>- to the C<sub>9</sub>-methyl group of retinal, which is visualized in Fig. 6.3.

The steric clash of the isomerized retinal with W223 is thus in good agreement to the experimental observation. However, longer timescales are needed to see any related changes, like a helix movement or deformation and is addressed to section 6.4.1.

Takemoto *et al.*<sup>208</sup> already suggested a steric clash between the C<sub>13</sub>-methyl group of retinal and the indol ring of W223 as obtained by MM MD simulation on the C1C2 structure. This subsequently induced a movement of the cytoplasmic half of helix G. Ardevol *et al.*<sup>13</sup> found also the movement of the indol ring of W223 from the C<sub>13</sub>- to the C<sub>9</sub>-methyl group of retinal, but assigned it to the early P<sub>2</sub> state. However, this might report a too late steric clash, since experimental studies<sup>134,135,161,183</sup> highlighted the fast formation of the amide I band at 1664 cm<sup>-1</sup>. The time when this steric clash happens, might be important for further processes, as discussed in section 6.4.1 and 6.4.2.

## DC Gate

The DC gate shows its importance in the channel functionality, since mutation in this part leads to drastic changes in the channel kinetics. FTIR experimental results even highlighted the predicted H-bond changes in the DC gate as observed by shifts of vibrational bands in the carboxylic region of D156.<sup>135</sup>

Hence, the H-bond percentage over all QM/MM MD trajectories for both homology model and X-ray structure are calculated and compared to the dark state, see Tab. 6.5. No significant changes are obtained compared to the dark state model independent of the QM region used for the homology model of system LW. The values of the largest QM region (QM-E90) are shown in Tab. 6.5. The detailed listing of the percentage of the H-bond pattern of the DC gate can be found in Tab. A.9 in the appendix.

**Table 6.5** Percentage of the H-bond (%) occurring in the DC gate averaged over all QM/MM MD trajectories, respectively. In the case of system HW, the values originated from the small QM region (QM) due to inconsistency in QM-TS. System LW shows for all QM region similar results, here shown QM-E90. In the case of the X-ray structure, the values averaged over all structural motifs are shown.

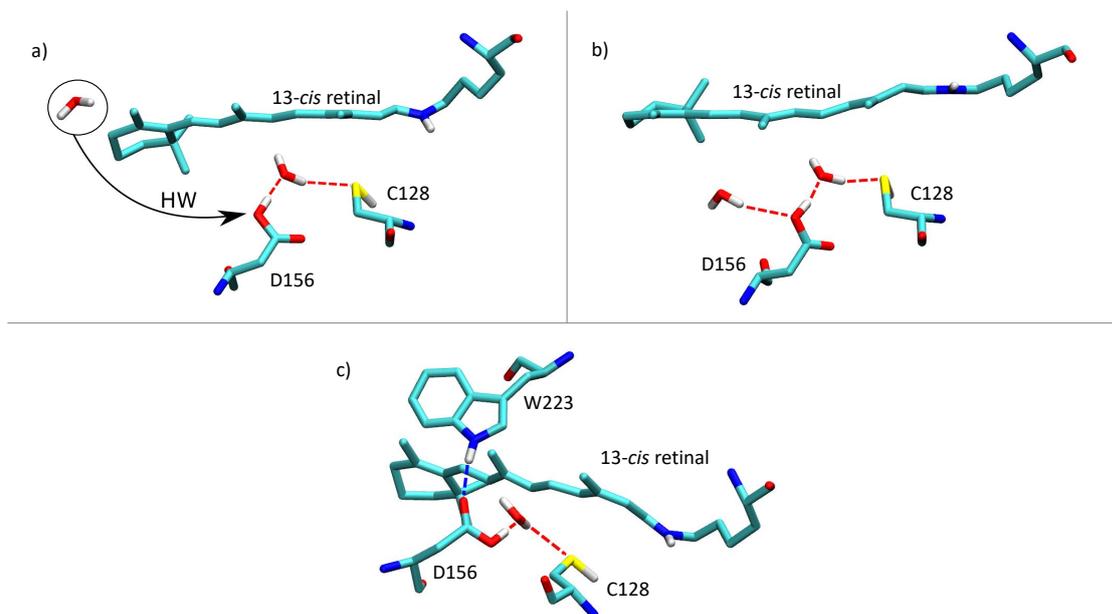
Model	structure motif	(D156)O···H <sub>2</sub> O	D156···H <sub>2</sub> O···C128
homology model	dark state	98.8 <sup>a)</sup>	32.6 <sup>a)</sup>
	system HW	83.5	25.4
	system LW	94.8	34.0
X-ray		95.9	32.0

<sup>a)</sup> QM/MM MD trajectories of Ref. 73, values calculated by F. Wolff

Differences are obtained in the QM/MM MD simulations of system HW due to the higher

water density. A nearby located water molecule moves towards the DC gate and forms a bond with D156, so that D156 is forming two H-bonds with the water molecules, see Fig. 6.4. Regarding the H-bond percentage in Tab. 6.5, a reduction of about 10 % is obtained, which indicates a weakening of the water-bridge DC gate and might originate from the second water molecule.

In the case of the X-ray structure, the same observation is found as obtained with system LW, i.e. no significant change of the H-bond percentage of the water-bridged DC gate compared to the dark state. However, the movement of the indol ring of W223 from the C<sub>13</sub>- to the C<sub>9</sub>-methyl group of retinal lead to the formation of a second H-bond between the indol ring (NH) of W223 and D156, see Fig. 6.4. Dependent on the time, when W223 moves towards the C<sub>9</sub>-methyl group of retinal, the hydrogen bond is found in about ~ 10-40 % of the snapshots of the QM/MM MD simulations. Further analysis are given in section 6.7, where the vibrational frequencies of several DC gate motifs are calculated to directly compare with experimental results. The other previous computational studies<sup>13,37,208</sup> on the P<sub>1</sub> state did not analyze the H-bond pattern of the DC gate. However, a detailed analysis of the H-bond pattern in the DC gate is important, since the experimental suggested substates<sup>135</sup> differ mainly in the DC gate H-bond pattern.

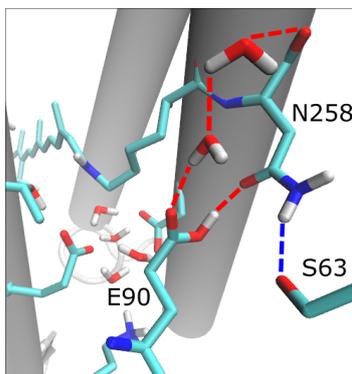


**Figure 6.4** H-bond DC gate pattern obtained in the QM/MM MD simulation of P<sub>1a</sub>. a) In system LW as well as in the X-ray pattern a stable water-bridged DC gate is obtained. The displayed second water molecule is only found in system HW. b) This second water molecule moves to the DC gate during the simulations forming a second H-bond. c) The movement of the indol ring W223 from the C<sub>13</sub>- to the C<sub>9</sub>-methyl group of retinal lead to the formation of an additional H-bond to D156 as found in the X-ray QM/MM MD simulations.

### Central Gate

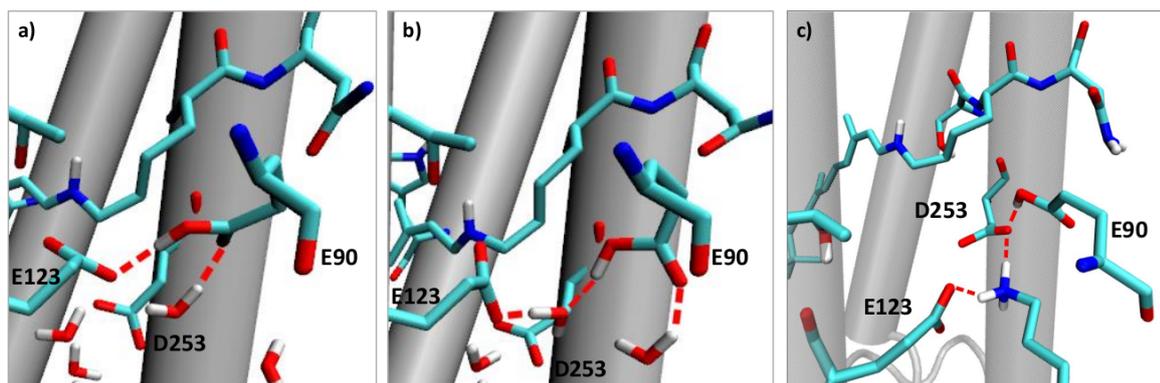
The central gate involves the residues E90, N258 and S63. Mutations in the central gate lead to changes in the ion selectivity. H-bond changes might occur as well in the central gate, suggested by FTIR results, which obtained a blue shift of the  $1717\text{ cm}^{-1}$  band to  $1728\text{ cm}^{-1}$ . This was assigned to H-bond changes of the carboxylic group of E90.<sup>133,134</sup>

The main difference between the homology model and the X-ray structure is found in the central gate in the different orientation of E90. Fig. 6.5 shows the typical H-bond network of the central gate in the homology model as presented in the dark state. E90 is oriented upward towards the cytoplasmic side and forms one direct H-bond to the sidechain of N258 and a two water-bridged H-bond to the backbone of N258. No changes are observed in the QM/MM MD simulations of the early P<sub>1a</sub> state for both systems HW and LW. The percentage of about  $\sim 99\%$  indicates a strong H-bond between E90 and N258 sidechain. The H-bond percentage of the central gate are tabulated in the appendix, see Tab. A.10. Comparing the different QM regions, the inclusion of the central gate into the QM region leads to a stronger H-bond of N258 with residue S63. Differences are found in the H-bond between the water molecule and E90. Here, the percentage is reduced about a factor of one half with a larger QM region including the central gate. However, E90 is still strong bonded via N258.



**Figure 6.5** E90 motif found in the homology model. E90 is upward oriented towards the cytoplasmic side and hydrogen bonded to one water molecule and N258.

The downward orientation of the X-ray structure leads to a different picture. In the experimentally obtained X-ray structure,<sup>217</sup> E90 is bonded in the dark state via a water molecule to E123 and D253, the counterions of the RSBH<sup>+</sup>, as well as to K93. In the QM/MM MD simulations of the early P<sub>1a</sub> state three different E90 H-bond pattern are obtained and showed in Fig. 6.6. In two of the 16 QM/MM MD simulations E90 flips also upward. The remaining QM/MM MD simulations show either a water-bridged H-bond to E123 and D253 or a direct bond to E123 or D253. A direct bond to K93 is also found, when K93 is moved upward towards E123 and D253, see Fig. 6.6 c.



**Figure 6.6** E90 orientation and H-bond pattern observed by the QM/MM MD simulation of the early P<sub>1a</sub> state of the X-ray structure. a) E90 forms a direct H-bond to E123. b) E90 forms a water-bridged bond to E123 (or D253). c) E90 forms a direct H-bond to D253.

Furthermore, the percentage of the direct H-bond of E90 to E123 or D253 is computed, since these H-bond pattern are easier to compare with the dark state motif. The reason is, that the water molecules can exchange and make it more difficult to investigate the H-bond percentage of E90 bonded to a water molecule over the trajectories.

The dark state QM/MM MD trajectories generated by F. Wolff are used for the calculation of the percentage of the E90 H-bond pattern. The dark state shows a direct H-bond between E90 and E123, but no bond to D253, see Tab. 6.6.

In the QM/MM MD simulations of the early P<sub>1a</sub> state an increase of the direct H-bond of E90 to E123 by up to maximum of  $\sim 40\%$  is obtained, which might indicate the blue shift of the  $\nu_{COOH}$  band assigned to E90.<sup>133,134</sup>

**Table 6.6** Percentage of the H-bond pattern (%) of E90 bonded to E123 or D253 occurring in the QM/MM MD simulations of the X-ray structure. The values are averaged over all QM/MM trajectories, respectively. The QM region belongs to QM-E90 for both dark state and P<sub>1a</sub>.

Model	structural motif of the dark state	(E90)OH $\cdots$ O(E123)	(E90)OH $\cdots$ O(D253)
X-ray	dark state	51.0	0.0
	E123 <sup>down</sup> -RSBH <sup>+</sup> $\cdots$ -O(D253)	74.8	0.0
	E123 <sup>down</sup> -RSBH <sup>+</sup> $\cdots$ H <sub>2</sub> O	89.9	0.0
	E123 <sup>up</sup> -RSBH <sup>+</sup> $\cdots$ H <sub>2</sub> O	43.5 <sup>a)</sup>	98.9 <sup>a)</sup>
	E123 <sup>up</sup> -RSBH <sup>+</sup> $\cdots$ -O(E123)	58.5	0.0

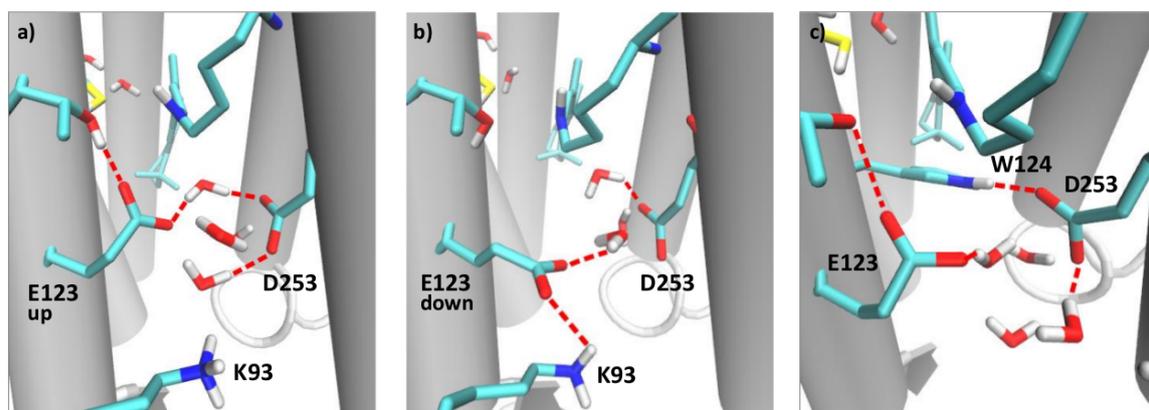
<sup>a)</sup> different motifs in the QM/MM MD simulations obtained, either E123 or D253 is dominated.

### E123 and D253

The isomerization of retinal leads to the disruption of the H-bond between RSBH<sup>+</sup> H-bond and the counterions E123 and D253. Hence, another H-bond pattern of both counterions is expected.

In the dark state, E123 showed two orientation: (i) an upward orientation towards the cytoplasmic side and hydrogen bonded to T127 or the RSBH<sup>+</sup>; (ii) a downward orientation towards the extracellular side and hydrogen bonded to water molecules located in the vicinity of the RSBH<sup>+</sup>. With about ~76 % the E123 upward orientation<sup>73</sup> was preferred as obtained by measurements on QM/MM MD trajectories defining a geometrical criteria, see below. The upward orientation was also confirmed energetically by subsequent well-tempered metadynamic simulations.<sup>73</sup>

The QM/MM MD simulations of the early P<sub>1a</sub> state of the homology model and the X-ray structure show the same E123 and D253 H-bond patterns, see Fig. 6.7. The H-bond pattern of D253 shows in general two motifs: (i) either hydrogen bonded to the water molecules between E123 and D253 or (ii) to the nearby located residue W124, see Fig. 6.7. The latter H-bond pattern is however less presented with ~ 2-7%, found in both homology model as well as in the X-ray structure. This indicates, that D253 is preferred bonded to the water molecules located between E123 and D253.



**Figure 6.7** E123 and D253 orientation and H-bond pattern found in the homology model and X-ray structure as obtained by the QM/MM MD simulation. a) E123 is oriented upward towards the cytoplasmic side and hydrogen bonded to T127 and a water molecule. b) E123 is oriented downward towards the extracellular side and hydrogen bonded to water molecules or as in this case to K93. c) D253 is hydrogen bonded to W124 or to water molecules located between E123 and D253.

In the homology model and the X-ray structure, E123 is also either upward or downward oriented. E123 can either forms a H-bond to T127 or the water molecules between E123 and D253. The geometrical criteria for the E123 orientation is defined by the dihedral angle around the C<sub>β</sub>-C<sub>γ</sub> bond, i.e. upward/downward for an angle above/below 240°, details see Ref. 73.

The same procedure is used here to identify the orientation pattern of E123 in the QM/MM MD simulations of the early P<sub>1a</sub> state of both homology model and X-ray structure. The results are summarized in Tab. 6.7. Since the different QM region used for the homology model show no differences, only the results of the large QM region (QM-E90) of system LW are shown here for comparison with the X-ray structure.

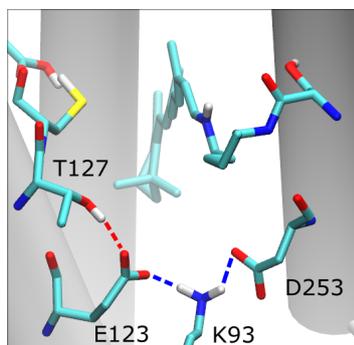
The results in Tab. 6.7 shows, that E123 is mainly upward oriented ( $\sim 90\%$ ), thus stabilized through the H-bond to T127. The percentage of E123 upward orientation is about 10-20% higher than obtained in the dark state, which results by the isomerized retinal since the H-bond to RSBH<sup>+</sup> is not preserved. Hence, a few flexible E123 behavior is obtained for the early P<sub>1a</sub> state.

**Table 6.7** Percentage (%) of the E123 orientation obtained during the QM/MM MD simulations of both homology model and X-ray. All values are averaged over all QM/MM MD trajectories, respectively.

Model	structural motif of the dark state	QM region	E123 upward	E123 downward
homology model	system LW <sup>a)</sup>	QM-E90	97.9	2.1
X-ray	E123 <sup>down</sup> -RSBH <sup>+</sup> ... <sup>-</sup> O(D253)	QM-E90	80.3	19.7
	E123 <sup>down</sup> -RSBH <sup>+</sup> ... H <sub>2</sub> O		88.7	11.3
	E123 <sup>up</sup> -RSBH <sup>+</sup> ... H <sub>2</sub> O		91.7	8.3
	E123 <sup>up</sup> -RSBH <sup>+</sup> ... <sup>-</sup> O(E123)		99.2	0.8

<sup>a)</sup>E123<sup>up</sup>-RSBH<sup>+</sup> ...<sup>-</sup>O(E123)

In the X-ray structure, E123 as well as D253 can be also directly bonded to E90 as described in the previous section. In this case, E123 and D253 are oriented parallel to each other, otherwise there are shifted to each other. The parallel orientation of E123 and D253 is also formed when K93 flips upward towards the cytoplasmic side, see Fig. 6.8. This pattern is only found in X-ray QM/MM MD simulations.



**Figure 6.8** In some QM/MM MD simulations of the X-ray structure K93 flips upward towards the cytoplasmic side forming an H-bond to E123 and D253.

## 6.4 MM MD Simulations of P<sub>1</sub>

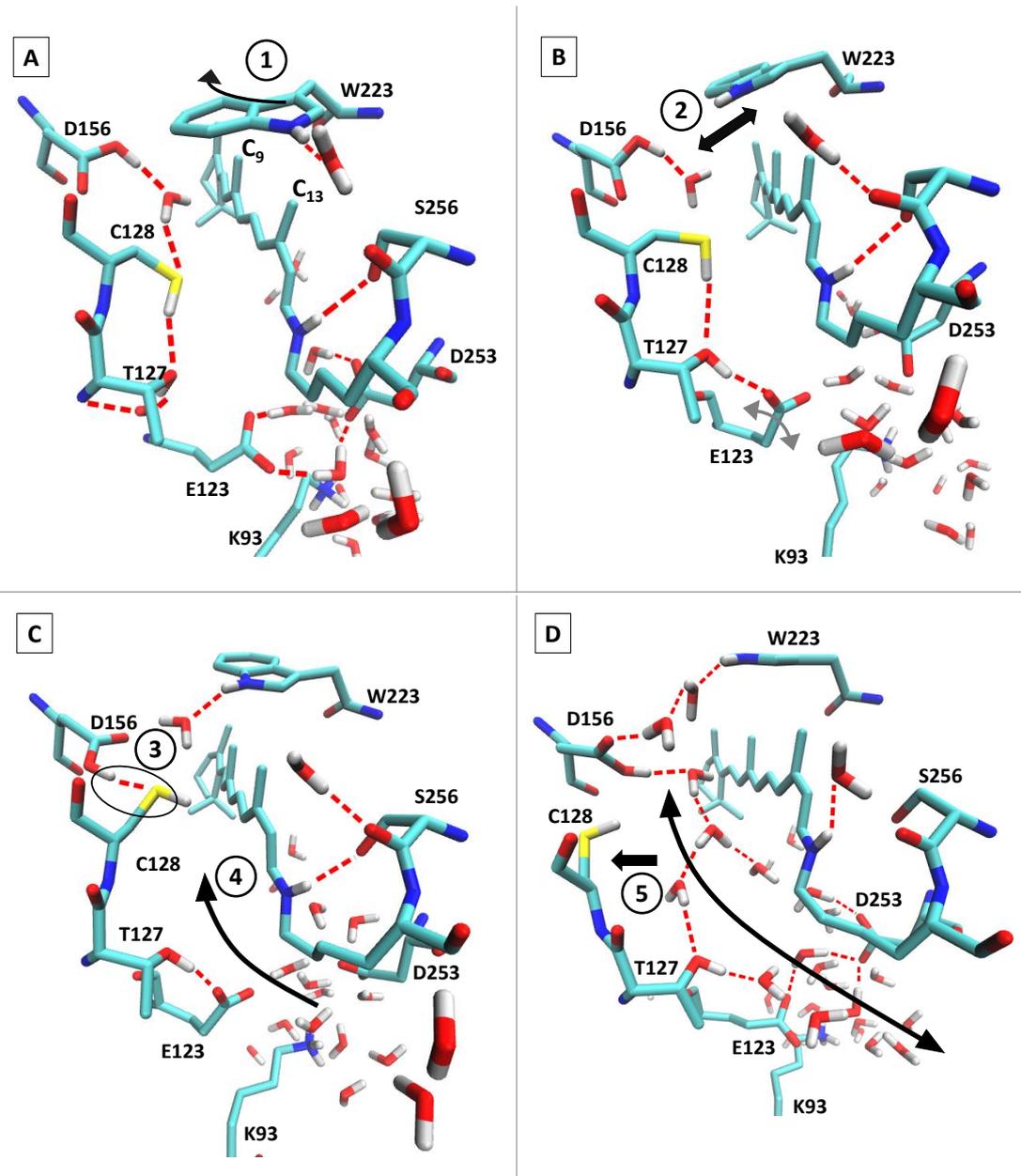
The lifetime of the first intermediate state P<sub>1</sub> is about 400 ns long. To sample additionally several substates, MM MD simulations are performed for both homology models, system HW and LW, as well as for the X-ray structure. The MM MD trajectories are analyzed and the results are summarized in the following subsections.

### 6.4.1 Pre-Gating Process

The first intermediate state P<sub>1</sub> is involved in the pre-gating process, i.e. the preparation for the deprotonation of the RSBH<sup>+</sup> and the subsequent preparation for channel opening. The analysis of the MM MD simulations for the homology model and the X-ray structure reveals a process, which includes changes in the DC gate and also the pre-formation of a possible water pore. This observed pre-gating process is summarized in Fig. 6.9 and described in the following. Subsequently, the structural differences are described found within the several MM MD simulations, which lead to some deviations of the process shown in Fig. 6.9.

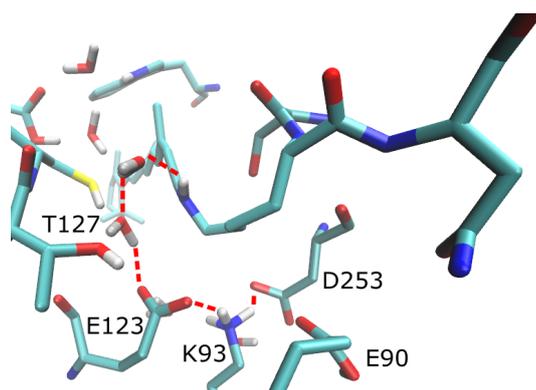
In the first nanoseconds of the MM MD simulations the indol ring of W223 moves from the C<sub>13</sub>- to the C<sub>9</sub>-methyl group of the 13-*cis* retinal. This observation is also found in the QM/MM MD simulations with the large QM region (QM-E90) of both homology model and X-ray structure. Analogous to the X-ray QM/MM MD simulations, the indol ring of W223 forms a direct or water bridged bond to D156 in some MM MD simulations.

A turn of the indol ring of W223 towards the DC gate is found after its displacement from the C<sub>13</sub>- to the C<sub>9</sub>-methyl group of the 13-*cis* retinal. This turn pushes D156 from its previous position away, so that D156 moves closer to C128. The steric clash with the DC gate leads then to the formation of a direct bond between D156 and C128, which in turn forms a cavity between C128 and retinal. This cavity makes it possible for water molecules to influx, which subsequently pushes C128 from the retinal away. In some MM MD simulations this leads to a highly twisted retinal configuration, see Fig. 6.11, i.e. the C<sub>13</sub>-methyl group is turning towards C128. This turn however is only found in the homology model of system HW and in one trajectory of system LW. In one MM MD simulations of the X-ray structure and two of the homology model a back relaxation of the DC gate is found, however the higher water amount is still present. Thus, the DC gate is either directly bonded or again water-bridged. This might indicates an interstep in this pre-gating process for the preparation of channel opening.



**Figure 6.9** The main pre-gating process as obtained by MM MD simulations of the homology model and the X-ray structure. **A:** Step ① describes the movement of the indol ring of W223 from the C<sub>13</sub>- to the C<sub>9</sub>-methyl group of 13-*cis* retinal. **B:** This displacement leads to a movement of the indol ring of W223 towards the DC gate, which subsequently pushes D156 away (step ②). The gray error on E123 indicates, that in the MM MD trajectories E123 occurs in two orientations, i.e. upward and downward. **C:** The steric clash between the indol ring of W223 and the DC gate leads to step ③: the formation of a direct hydrogen bonded DC gate. The consequences are a cavity formed between C128 and retinal (step ④). **D:** In step ⑤ water is influxed, which subsequently pushes C128 from the retinal away.

The higher water density obtained during the MM MD simulations resulted within the pre-gating process, leads to a RSBH<sup>+</sup> H-bond pattern, as shown in Fig. 6.10. This water-bridged hydrogen bond to E123 suggests, that the subsequent deprotonation might occur via a water molecule. Thus, the acceptor is addressed to E123 and might go via K93 towards D253, since D253 is suggested to be the proton acceptor of the RSBH<sup>+</sup> as obtained by FTIR measurements.<sup>134</sup> The motif K93 bridging E123 and D253, is observed by the X-ray structure but also in system HW. This deprotonation pathway would be similar to bR, where the proton acceptor of the RSBH<sup>+</sup> is the amino acid D85.<sup>25, 68, 136</sup>

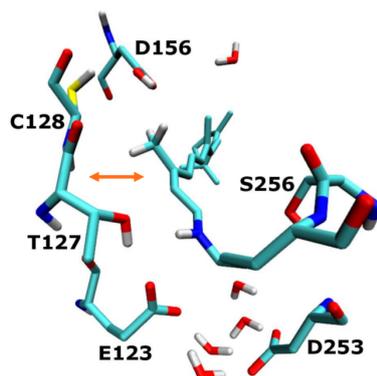


**Figure 6.10** RSBH<sup>+</sup> is water-bridged to E123 after water influx due to the movement of the DC gate.

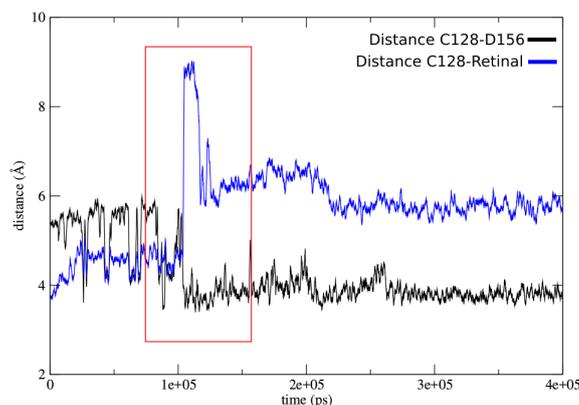
## 6.4.2 The Retinal Binding Pocket

### Retinal Configuration

As described above, a turn of the C<sub>13</sub>-methyl group of retinal towards C128 is found in the MM MD simulations. Cheng *et al.*<sup>37</sup> found also a highly twisted 13-*cis* retinal configuration when performing MM MD simulations on the early P<sub>2</sub> state, i.e. when the RSB is already deprotonated. They concluded, that the turn of the C<sub>13</sub>-methyl group of retinal towards C128, pushes C128 away. The finding as well as the conclusion are however, in contrast to our results. As described above (Fig. 6.9), the highly twisted retinal is also found, but occurs in the P<sub>1</sub> state and is not the reason for the displacement of C128. To visualize this, one trajectory of the MM MD simulations is chosen, where the turn of the C<sub>13</sub>-methyl group of retinal is found. The distances between the sulfur atom of C128 and the oxygen atom of D156, i.e. the oxygen involved in the H-bond formation, are measured over the trajectory. Additionally, the distance between C128 and retinal is also measured, where the sulfur atom of C128 is chosen as well as the C<sub>13</sub> atom of retinal. These distances are plotted together in Fig. A.4.



**Figure 6.11** The direct hydrogen bond of D156 to C128 leads to a cavity between C128 and retinal. Water can influx and pushes C128 from the retinal away. Subsequently, a turn of the 13-*cis* retinal towards C128 is found.



**Figure 6.12** Distances between (D156)O–S(C128) (black) and (C128)S–C<sub>13</sub>(retinal) (blue). The values are averaged for each points (10) to reduce the fluctuations. The red bordered area shows the abrupt change when C128 moves away from the retinal. The distance decreases again, when the C<sub>13</sub>-methyl group of retinal turns towards C128.

The distances between the DC gate, i.e. between D156 and C128 changes when a direct bond is formed. At the beginning the water-bridged DC gate shows a distance of (D156)O–S(C128) around 5.5 Å, while the steric clash with W223 and D156 leads to the formation of a direct bond accompanied with a reduced distance of (D156)O–S(C128) to about 3.5 Å.

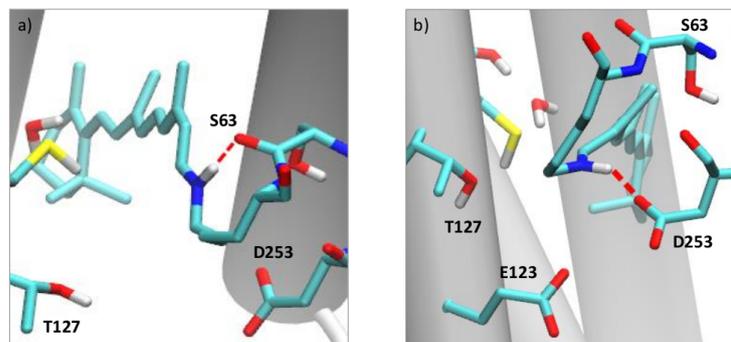
As described above, the water influx pushes C128 from the retinal away, which is shown in the distance of (C128)S–C<sub>13</sub>(retinal) over the trajectory. At the beginning, the distance is about 4.5 Å, and increases to 9 Å when C128 is pushed from the retinal away and is decreased due to the turn of the C<sub>13</sub>-methyl group of retinal. The turn of the C<sub>13</sub>-methyl group of retinal is visualized in Fig. A.4 in the appendix by the measurement of the dihedral angle C<sub>10</sub>–C<sub>11</sub>–C<sub>12</sub>–C<sub>13</sub>. The twist of the retinal occurs at the same time, when the distance of

(C128)S–C<sub>13</sub>(retinal) decreases.

Even though, Cheng *et al.*<sup>37</sup> reported also the twist of the retinal, two complete different processes are obtained. Here, the twist is only found in the MM MD simulations of the homology model. Both findings are distinguished in (i) when the twist occur, i.e. in our case in P<sub>1</sub> and in Cheng *et al.*<sup>37</sup> in the early P<sub>2</sub> with the deprotonated RSB; (ii) the reason and the effect of the turn of the C<sub>13</sub>-methyl group of retinal, i.e. in our case, the consequence of the cavity between C128 and retinal and in Cheng *et al.*,<sup>37</sup> the turn of the C<sub>13</sub>-methyl group pushes C128 away. Both processes might be possible. Further investigation of the deprotonation of the RSB with subsequent simulations might clarify, whether this twist can be also found in P<sub>2</sub> or whether Cheng *et al.*<sup>37</sup> modeled the deprotonation too early due to less sampling. Furthermore, it can be analyzed, whether this twist occurs only in the homology model or also with the X-ray structure.

### Hydrogen Bonds of the RSBH<sup>+</sup>

The study of the QM/MM MD trajectories gives already an insight about the H-bond pattern of RSBH<sup>+</sup> in the early P<sub>1a</sub> state. The large QM region (QM-E90) reveals, that the RSBH<sup>+</sup> is about 20 % hydrogen bonded to T127. The small QM region, i.e. where no nearby residue of the RSBH<sup>+</sup> is involved shows a preferred H-bond to the residue S256. This same observation is found in all MM MD simulations. Additional, another H-bond pattern of the RSBH<sup>+</sup> is found in all MM MD simulations of the homology model and the X-ray structure, see Fig. 6.13 a. The H-bond is formed between the RSBH<sup>+</sup> and the backbone of S256.



**Figure 6.13** In the MM MD simulations two H-bond pattern of the RSBH<sup>+</sup> are found. a) RSBH<sup>+</sup> bonded to the backbone of S256, found in all MM MD simulations of the homology model and the X-ray structure. b) RSBH<sup>+</sup> ··· O-(D253), only found in system HW.

This observation is in agreement to the QM/MM MD simulations investigating the influence of different QM regions. The force field neglects one H-bond pattern of the RSBH<sup>+</sup>, namely the hydrogen bond to T127. This can have influences on the further investigations. However, the analysis of the QM/MM MD simulations revealed a percentage of about 40 % of the

RSBH<sup>+</sup> ··· O-(S256) H-bond pattern with the small QM region, i.e. the exclusion of T127 and S256 from the QM region. This indicates a not strongly bonded RSBH<sup>+</sup>. It is known, that the QM description improves the hydrogen bond description, as described in chapter 3. Nevertheless, MM MD simulations are important to sample several substates.

Furthermore, another difference is found in the MM MD trajectories of the homology model of system HW showing a H-bond of the RSBH<sup>+</sup> to D253 (RSBH<sup>+</sup> ··· O-(D253)), while the 13-*cis* retinal is highly twisted, see Fig. 6.13 b. The time period of this H-bond pattern is different, spanning a time from about 9 ns up to 180 ns. The longer the RSBH<sup>+</sup> is hydrogen bonded to D253, the longer the retinal is highly twisted and no steric clash between the indol ring and the 13-*cis* retinal occurs. Hence, this H-bond pattern leads to a deviation of the previously described pre-gating process, since it lagged the time for the movement of the indol ring of W223. However, this seems to be in conflict with experimental results. They suggested an early steric clash of one amino acid due to the fast obtained band at  $\sim 1664\text{ cm}^{-1}$  in the amide I region.<sup>134, 135, 161, 183</sup>

Interestingly, Ardevol *et al.* reported also the H-bond pattern RSBH<sup>+</sup> ··· O-(D253), as already discussed in section 6.3.1. Additionally, they found the displacement of the indol ring of W223 from the C<sub>13</sub>-methyl group of retinal to the C<sub>9</sub>-methyl group in the early P<sub>2</sub> state. However, our extensive sampling shows, that on the one hand the H-bond pattern RSBH<sup>+</sup> to D253 is either a force field artefact or the effect of too less sampling, since this occurs only in system HW, the less equilibrated system. On the other hand, this H-bond pattern leads to a delayed steric clash with the indol ring W223, which is not in agreement with the experimental results, as described above. Furthermore, this H-bond pattern is not found in the X-ray structure, and might occur only in the C1C2 model.

### Central Gate, E123 and D253

The central gate in the MM MD simulations behaves differently dependent on the orientation of E90. The homology model and the X-ray structure lead to a different observations.

The homology model contains an upward oriented E90 configuration towards the cytoplasmic side. In both systems HW and LW, a downward flip of E90 in five of eight MM MD simulations is obtained. Subsequently, the same H-bond pattern is found in the QM/MM MD simulations of the X-ray structure. Hence, E90 is hydrogen bonded to water molecules between E123 and D253 or directly hydrogen bonded to E123 and D253. A difference is obtained in system HW, which shows the upward flip of K93 as also found in the QM/MM MD simulations of the X-ray structure, see Fig. 6.8. The consequences of the downward flip of E90 is described in the following subsection.

The mechanism of the pre-gating process as obtained by the MM MD simulations shows no significant contribution of the central gate, when E90 is oriented downward as presented in the X-ray structure. The analysis of the QM/MM MD simulations reveals the same H-bond

pattern of E90, as also found here. Interestingly, the residue K93 turns also upward, and thus forms a H-bond to E123 and D253 bridging these two residues. This occurs mainly at the beginning in all MM MD simulations.

Furthermore, in the MM MD simulations an upward and downward orientation of E123 is also found. The percentage of these orientation is given in Tab. 6.8, which is calculated in the same way as described in section 6.3.1.

**Table 6.8** Percentage (%) of the E123 orientation obtained during the MM MD simulations of the homology model and the X-ray structure. All values are averaged over all MM MD trajectories, respectively.

Model	trajectory	E123 upward	E123 downward
homology model, system HW	1	59.7	40.3
	2	25.2	74.8
	3	37.2	62.8
	4	28.9	61.1
homology model, system LW	1	62.9	37.1
	2	37.4	62.6
	3	65.6	34.4
	4	76.6	23.4
X-ray	1	42.7	57.3
	2	49.2	50.8
	3	23.2	76.3
	4	4.00	96.0

The MM MD simulations of the homology model show a preferred downward orientation of E123, when a downward flip of E90 occurs. The X-ray structure as shown in Tab. 6.8 obtains also mostly a downward orientation of E123. However, in the QM/MM MD simulations of the early P<sub>1a</sub> the upward orientation is dominated. This could indicate, that either in the pre-gating process the orientation of E123 is changed, or this is a force field artefact due to poorly described H-bond between T127 and E123. Further discussions are given in section 6.5.1.

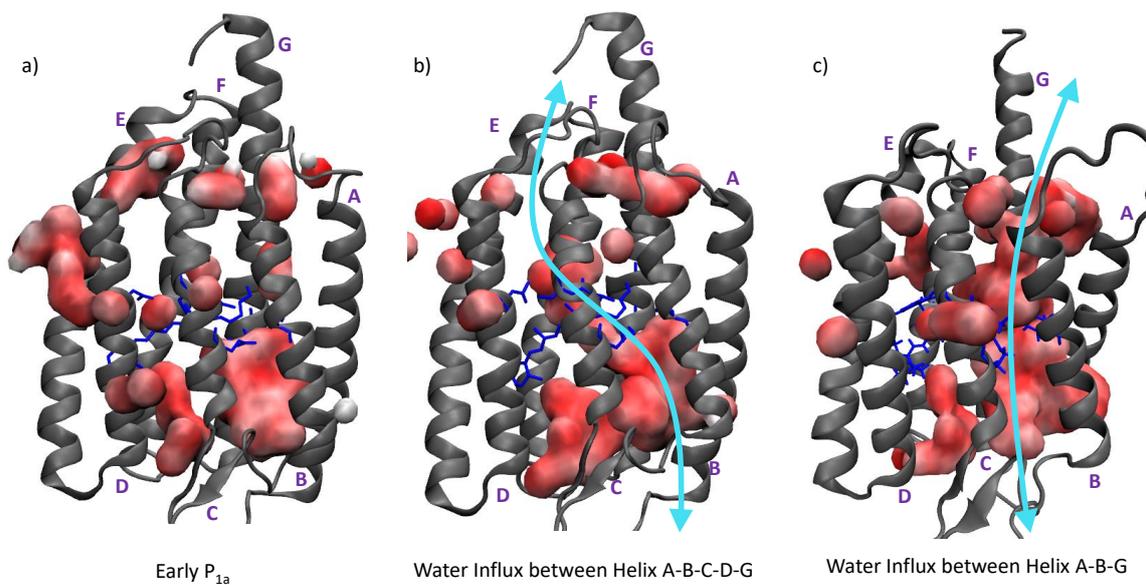
### 6.4.3 Helix Movements and Water Influx

The previously described structural and H-bond changes can lead to helix movements, deformations or possibly also to a water influx. These changes are suggested by experimental studies, which indicate that the formation of a larger pore might be necessary due to the cation conductivity in contrast to only proton pumps.<sup>134,135,161,183</sup> Furthermore, the formation or enlargement of water-filled pores inside the protein might be a prerequisite for cation permeation.<sup>132,134</sup>

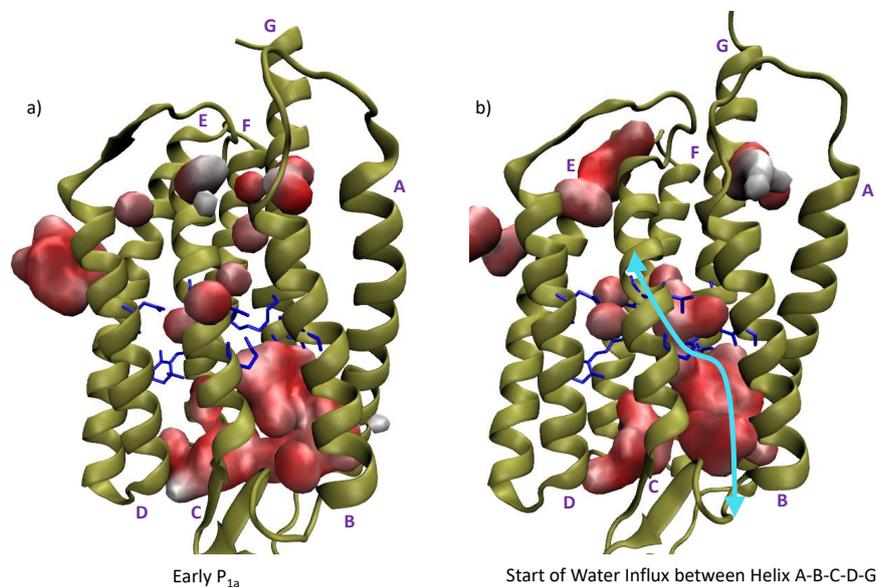
The RMSD of every helix is calculated for all MM MD trajectories and plotted for the homology model, here system LW and the X-ray structure. The plots are found in the appendix, see Fig. A.5-11. The comparison of the RMSD of the homology model with those of the X-ray structure reveals no significant differences except of helix C. Helix B, D and G show a RMSD in a range of 1.0-1.5 Å. A slightly higher range is obtained by helix A and E about 1.0-2.0 Å and even slightly higher by helix F. Helix C shows in the homology model a RMSD in a range of 1.0-1.5 Å, while the X-ray structure obtains a RMSD of only 0.5-1.0 Å.

The higher RMSD value for helix F results by the movement of the indol ring of W223, since W223 is located on helix F. The main difference is obtained by helix C, since the mechanism for the pre-gating process reflects changes of the orientation of C128 located on helix C. However, in the X-ray structure, C128 relaxes often back, and thus no significant changes are observed. This is accompanied with a slightly reduced water density, since few water molecules moving between retinal and the DC gate. The reason and consequences are discussed below. Additionally, one trajectory of the X-ray structure shows a higher RMSD, since in this trajectory an outward movement of D156 located on helix D is found. In general, nearly all helices move slightly, but no helix shows a high increase of RMSD. More sampling is necessary to see, whether an outward tilt will be already observed in the early pre-gating mechanism or will occur later after the deprotonation of the RSBH<sup>+</sup>.

The structural changes as described by the pre-gating process leads also to a subsequent water influx. Fig. 6.14 and Fig. 6.15 show the water distribution at the beginning of the simulations and at the end of the simulations for the homology model and the X-ray structure, respectively. Fig. 6.14 b shows the water influx as obtained when the DC gate is forming the cavity between C128 and retinal. This indicates a water pore formation between helix A-B-C-D-G, while the water pore on the extracellular side is located between helix A-B-G, and on the cytoplasmic side between C-D and maybe F and E as indicated by the blue arrow. Fig. 6.14 c shows a complete water pore formation between helix A-B-G. This is found by one trajectory of system LW, where E90 flips downward and induces this water influx. The water pore formation between helix A-B-G triggered by the flip of E90 was previously reported in computational studies based on the C1C2 structure and MM MD simulations.<sup>13,37,120</sup> However, this could be only the case in C1C2 or the homology model, but is not a possible process in the Chr2-WT, since E90 is already downward oriented in the dark state.



**Figure 6.14** Possible water pore formation as obtained by MM MD simulations of the homology model. a) Water distribution at the beginning of the simulation, found in P<sub>1a</sub>. b) After 400 ns simulation and changes occurring in the DC gate. c) Water pore induced by changes in the DC gate and additional flip of E90 downward towards the extracellular side.



**Figure 6.15** Possible water pore formation as obtained by MM MD simulations of the X-ray structure. a) Water distribution at the beginning of the simulation, found in P<sub>1a</sub>. b) After 400 ns simulation and changes occurring in the DC gate.

Fig. 6.15 shows the water distribution changes in the X-ray structure as obtained during the MM MD simulations. This is similar to the water distribution found in MM MD simulations of the homology model, see Fig. 6.14 b. However, the water density is less than in the homology model. This indicates, that either more water is necessary in the structure model, or the X-ray structure has to be sampled longer. Nevertheless, a higher water density seems to be important to make a water pore formation possible. This indicates, that the originally X-ray structure<sup>217</sup> does not account for the real structure as naturally occurred due to the few water density compared to the long sampled X-ray structure.<sup>232</sup>

The importance of the water amount was also discussed in a previous experimental study, suggesting that the formation of a water pore might be necessary for ion permeation.<sup>132</sup> They also discussed the role of the protein dynamic versus water dynamic in the preparation for channel opening. However, experimental findings can not determine, which one is more likely to be responsible for channel opening or if both are necessary. The dominated protein dynamic would accompany with primarily structural changes making water influx possible, while the dominated water dynamic describes it the other way around. Our study of the MM MD simulations reveals, that firstly structural changes happen with subsequent water influx, which in turn leads again to structural changes. Here, both protein and water dynamic are involved for the preparation of channel opening. This was also suggested, if large-scale protein changes are involved, from microseconds to seconds.<sup>60,132</sup>

Furthermore, experimental results revealed that the water influx temporally correlates ( $\tau = 10 \mu s$ ) with the deprotonation of the RSBH<sup>+</sup>.<sup>132,134</sup> Fig. 6.10 shows the water distribution near the RSBH<sup>+</sup> after water influx, which leads to a water bridge between RSBH<sup>+</sup> and E123. Additionally, this indicates, that the RSBH<sup>+</sup> might be protonated via a water molecule to E123. And the bridge via K93 might transfer the proton to D253, which is the experimentally suggested proton acceptor.

## 6.5 QM/MM MD Simulations of the Late Substate P<sub>1b</sub>

Experimental studies suggested a second substate called P<sub>1b</sub>, as obtained by time-resolved tunable quantum cascade laser (QCL) IR measurements.<sup>135</sup> This second substate might be accompanied with a transient weakening of the interhelical H-bond between D156 and C128 caused by a rotation or movement of helices D and C at around 200 ns after the formation of P<sub>1a</sub>. To investigate structural changes compared to the early P<sub>1a</sub> state, three different DC gate motifs (see Fig. 6.16) are chosen from the X-ray MM MD simulations. Since these motifs are also found in the homology model, the further QM/MM study is restricted to the X-ray structure. Hence, these motifs are used to perform QM/MM MD simulations.

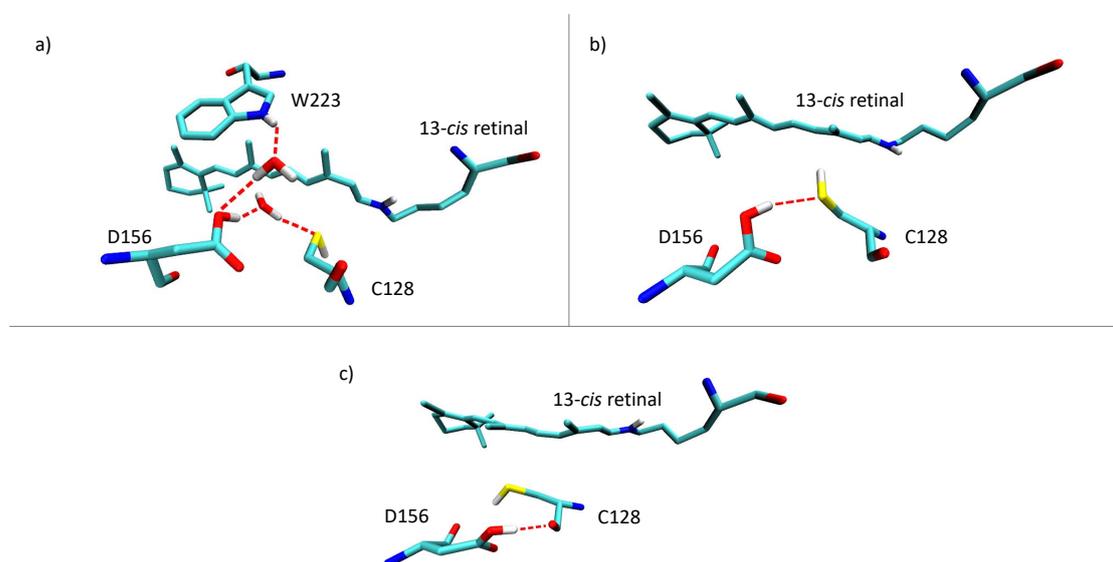
### 6.5.1 The Retinal Binding Pocket

#### Hydrogen Bond of the RSBH<sup>+</sup>

In contrast to the QM/MM MD simulation of the the early P<sub>1a</sub> state, no H-bond of the RSBH<sup>+</sup> to either T127 or S256 is found. The retinal is mostly planar and the RSBH<sup>+</sup> forms either no H-Bond or to a water molecule as shown in Fig. 6.10.

#### DC Gate

In the MM MD simulations different DC gate pattern are observed, see Fig. 6.16. These are distinguished through a direct bond of D156 to C128 or a water bridged bond. In some cases D156 forms a second bond directly to W223 or a water-bridged bond to W223, as obtained in the QM/MM MD simulations of the early P<sub>1a</sub> state.



**Figure 6.16** Different DC gate motifs as obtained by the X-ray MM MD simulations. a) water-bridged DC gate and a water-bridged H-bond between D156 and the indol ring of W223. b) direct DC gate. c) H-bond between D156 and the backbone of C128.

Motif a) in Fig. 6.16 describes a water-bridged DC gate H-bond, which stays stable during the whole simulation as obtained by the calculation of the H-bond percentage. The same percentage are observed as reported for the early P<sub>1a</sub> state. Motif b) in Fig. 6.16 shows a direct H-bond between D156 and C128, which remains stable during the QM/MM MD simulations. Additionally, in two out of five QM/MM MD simulation a direct H-bond to the indol ring W223 with about ~20% is found. Motif c) in Fig. 6.16 displays a H-bond of D156 to the backbone of C128, which remains also stable.

### Central Gate, E123 and D253

The QM/MM MD simulation of the late P<sub>1b</sub> state shows the same H-bond pattern of E90 than obtained by the QM/MM MD simulations of the early P<sub>1a</sub> state. In Tab. 6.9, the percentage of the H-bond pattern of E90 are displayed, analogous to the P<sub>1a</sub> state. In contrast to P<sub>1a</sub>, the directly hydrogen bonded E90 to either E123 or D253 obtains a high percentage about 91–100%. This indicates a strongly H-bond in contrast to the more flexible E90 H-bond pattern found in the P<sub>1a</sub> state, see section 6.3.1.

This significantly indicates a difference compared to the dark state H-bond pattern of E90, since in the dark state E90 is water-bridged hydrogen bonded to E123 or D253. Furthermore, this is in accordance to experimental studies, which suggested changes in the H-bond pattern of E90. FTIR measurements obtained a 11 cm<sup>-1</sup> blue shift of the  $\nu_{COOH}$  band (1717 cm<sup>-1</sup>(-)/1728 cm<sup>-1</sup>(+)) from the dark state to the first intermediate state. This band was assigned to E90, since this band disappeared when investigating the E90A mutant.<sup>133,134</sup>

Furthermore, within the central gate, the main difference between the X-ray structure and the homology model is found. The homology model containing the upward oriented E90 could not report any changes, since E90 remains strongly hydrogen bonded to N258. Even though, some MM MD simulations reveal a downward flip of E90 corresponding to H-bond changes, this is not possible in the X-ray structure due to already downward oriented E90 in the dark state. This downward flip was already reported by previous computational studies,<sup>13,37,120</sup> which was assigned to account for the H-bond changes in the central gate. However, this is only possible in the case of the C1C2 model, but not for Chr2-WT.

**Table 6.9** Percentage of the H-bond pattern (%) of E90, which is hydrogen bonded to E123 or D253. The values are averaged over all QM/MM MD trajectories, respectively. The QM region belongs to QM-E90.

Model	structural DC gate motif	(E90)OH...O(E123)	(E90)OH...O(D253)
	X-ray	D156...H <sub>2</sub> O...C128	0.0
	(D156)O...S(C128)	99.8	0.0
	(D156)O...O(C128)	0.74	90.8

The orientation of E123 is also investigated in the QM/MM MD simulations of the late P<sub>1b</sub> state. The QM/MM MD simulations of the X-ray structure of the early P<sub>1a</sub> state shows a preferred E123 upward orientation towards the cytoplasmic side with about 80-99%. However, during the MM MD simulations the downward orientation of E123 is found to be more dominated. Tab. 6.10 displays, that the E123 upward orientation is preferred during the QM/MM MD simulations. In the upward orientation E123 remains hydrogen bonded to T127. This motif is also obtained for the early P<sub>1a</sub> state. Due to the water influx observed by the MM MD simulation, another motif is found, showing a water-bridged E123 to T127 as shown

in Fig. 6.10. The different results obtained by MM and QM/MM MD simulations indicate, that the H-bond between T127 and E123 is stabilized by QM description.

D253 remains hydrogen bonded to either water molecules between E123 and D253 or E90. Only in one out of the 15 QM/MM MD simulations D253 is hydrogen bonded to W124 with about 32 % as also found in the QM/MM MD simulations of the early P<sub>1a</sub> state. Both, E123 and D253 remain bridged via K93. This motif originates from the MM MD simulations, see Fig. 6.8 and suggests to act as a proton transfer unit involving the deprotonation of the RSBH<sup>+</sup> as described in the previous section.

**Table 6.10** Percentage (%) of the E123 orientation. All values are averaged over all QM/MM MD trajectories, respectively.

Model	structural DC gate motif	E123 upward	E123 downward
X-ray	D156···H <sub>2</sub> O···C128	90.6	9.4
	(D156)O···S(C128)	72.8	27.2
	(D156)O···O(C128)	80.8	19.2

<sup>a)</sup>E123<sup>up</sup>-RSBH<sup>+</sup>···<sup>-</sup>O(E123)

## 6.6 Absorption Spectra of P<sub>1</sub>

The main characterization of the different intermediate states of the ChR2-WT photocycle is observed by UV/Vis spectroscopy. The extensive study of the dark state absorption spectrum in the case of the homology model revealed a good agreement to the experimental observed results, when investigating different structural motifs and comparison to the well studied bR absorption maximum.<sup>73</sup> The extension using the X-ray structure<sup>232</sup> obtained no significant differences compared to the homology model. Hence, it was concluded that the different E90 orientations do not effect significant changes on the absorption maximum.

The first intermediate state shows experimentally a red-shift of about 30 nm (0.16 eV).<sup>134</sup> As discussed in chapter 4 and 5, the absorption spectrum is mainly influenced by the geometry of the retinal configuration and the surrounded binding pocket.

Several structural motifs are obtained in the QM/MM as well as MM MD study on the first intermediate state P<sub>1</sub>. To investigate whether these motifs lead to the same or different shifts compared to the dark state, for all QM/MM MD simulations of the early P<sub>1a</sub> and late P<sub>1b</sub> state, the excitation energies are computed using OM2/MRCI. Thus, the excitation energies are computed of 1000 snapshots, which are obtained by the 1 ns long QM/MM MD trajectories, respectively. The excitation energies are weighted by the oscillator strength. A Gaussian distribution of all QM/MM MD trajectories is observed, respectively. The absorption maxima are obtained with a fit using a Gaussian function and displayed in Tab. 6.11.

**Table 6.11** Absorption maxima (eV) of ChR2-WT of the early P<sub>1a</sub> and late P<sub>1b</sub> obtained by the computation of excitation energies on sampled structures by QM/MM MD simulation with subsequent Gaussian fit of the histograms. All excitation energies are weighted by oscillator strengths.

Model	structural motif	QM region	P <sub>1</sub>	P <sub>1</sub> - DS
homology model <sup>b</sup>	system HW <sup>a</sup>	QM	2.96	-0.25
		QM-TS	2.86	-0.35
	system LW <sup>a</sup>	QM	2.97	-0.24
		QM-TS	2.98	-0.23
		QM-DC	2.98	-0.21 <sup>c</sup>
X-ray <sup>d</sup> , P <sub>1a</sub>	E123 <sup>down</sup> -RSBH <sup>+</sup> ... <sup>-</sup> O(D253)	QM-E90	2.93	-0.30
	E123 <sup>down</sup> -RSBH <sup>+</sup> ...H <sub>2</sub> O		2.89	-0.34
	E123 <sup>up</sup> -RSBH <sup>+</sup> ...H <sub>2</sub> O		2.96	-0.27
	E123 <sup>up</sup> -RSBH <sup>+</sup> ... <sup>-</sup> O(E123)		3.03	-0.20
P <sub>1b</sub>	D156...H <sub>2</sub> O...C128		2.89	-0.34
	(D156)O...S(C128)		3.00	-0.23
	(D156)O...O(C128)		2.98	-0.25

<sup>a</sup>E123<sup>up</sup>-RSBH<sup>+</sup>...<sup>-</sup>O(E123)

<sup>b</sup>dark state (DS) absorption maximum 3.21 eV<sup>73</sup>

<sup>c</sup> DS absorption maximum 3.21 eV, using the QM-region QM-DC<sup>232</sup>

<sup>d</sup>DS absorption maximum 3.23 eV<sup>232</sup>

The absorption maxima obtained for the homology model and the X-ray structure are in the same range, around 2.89-3.03 eV for the large QM-region (QM-E90). In the case of the homology model, using different QM regions, the values are also in the same region except of the QM region QM-TS of system HW, showing a higher red shift.

In general, the absorption maxima obtained for the first intermediate state P<sub>1</sub> for both, homology model and X-ray, show a red-shift compared to the dark state absorption maximum. This is qualitatively in agreement with the experimental obtained shift. The experimental red shift is about 0.16 eV, while the obtained shift by the OM2/MRCI calculations is slightly higher in a range of 0.20-0.34 eV. This deviation originates from several influencing factors.

The study of the absorption maxima of the dark state obtained also a range of 3.15-3.23 eV depending on the structural motifs. Hence, the absorption maxima area is similar to those of the P<sub>1</sub> state and reflects a high flexibility of the ChR2-WT structures in contrast to other rhodopsins like bR,<sup>82</sup> which are few flexible.

The isomerization study revealed the most probably pathway for the retinal isomerization, when started from H-bond pattern RSBH<sup>+</sup>...<sup>-</sup>O(E123) in the dark state. Taking this fact into account, the computational obtained shifts of 0.20 eV matches the best to the experimental

obtained shift (0.16 eV) between the dark state and P<sub>1</sub>.

The comparison of the absorption maxima for the P<sub>1a</sub> and P<sub>1b</sub> state shows, that both are found in the same range. The analysis of both QM/MM MD trajectories reveals however differences in structural changes and H-bond network changes. This is in agreement to the experimental results, which obtained for the P<sub>1</sub> substates no difference through UV/Vis spectroscopy. The differences resulted by FTIR measurements<sup>135</sup> are discussed in the next section.

## 6.7 Characterization of P<sub>1a</sub> and P<sub>1b</sub>: Vibrational Frequencies

The first intermediate state P<sub>1</sub> spans a time period of about 400 ns. In this time period several P<sub>1</sub> substates can appear, which might be spectroscopic silent and thus cannot be identified by experimental studies. In general, the computation of vibrational frequencies goes hand in hand with the experimental study, so that a direct comparison is possible. Especially the region of 1710-1760 cm<sup>-1</sup> where  $\nu_{COOH}$  bands of the glutamatic or aspartic acids can be found, is free from overlap with other residues.<sup>44</sup>

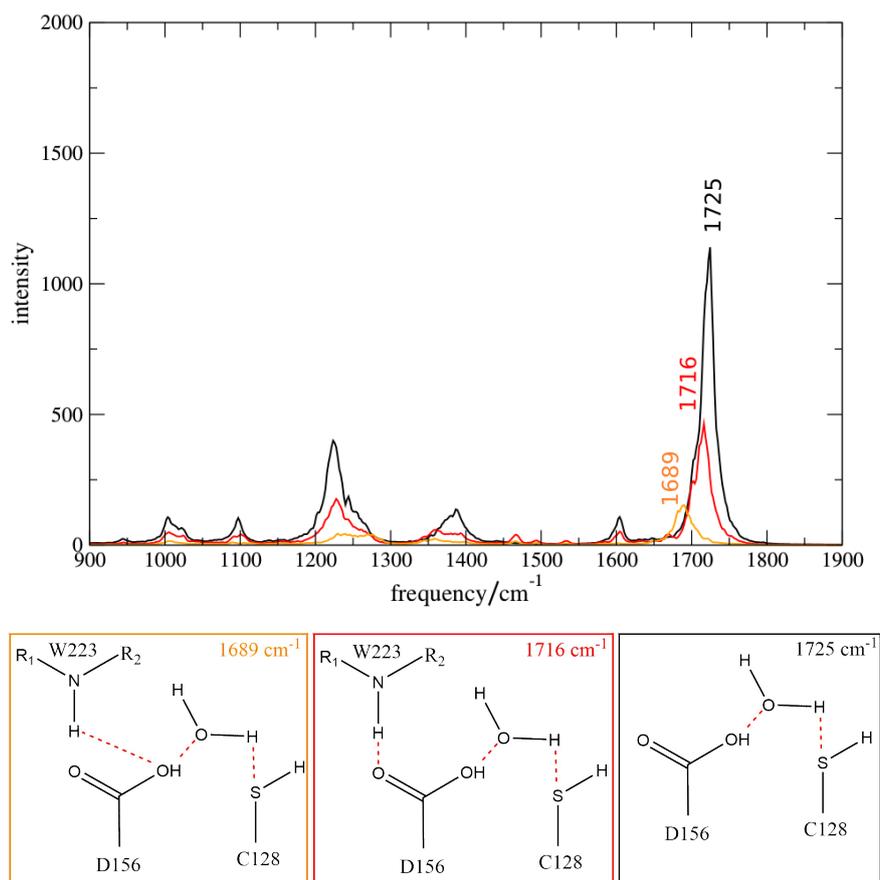
FTIR studies suggested, that the P<sub>1</sub> is divided into two substates, which are not distinguish by UV/Vis spectroscopy. The definition of the two substates is based on the different vibrational bands assigned to the DC gate. In the carboxylic region a blue shift of the 1737 cm<sup>-1</sup> band in the dark state to 1745 cm<sup>-1</sup> in P<sub>1</sub> was obtained and assigned to the weakening of the H-bond of D156. This observation differs from other rhodopsins, since e.g. in bR a red-shift of this band with smaller intensity is obtained for the first intermediate state (1740(-)/1733(+) cm<sup>-1</sup>).<sup>28,133,153,180</sup> Furthermore, in ChR2-WT a broad band at 1760 cm<sup>-1</sup> is found in P1 originating additional from the C=O stretching of D156. This band was then assigned to the late P<sub>1b</sub> state and suggested a transient weakening of the interhelical H-bond between D156 and C128 caused by a rotation or movement of helices D and C at around 200 ns after the formation of P<sub>1a</sub>.<sup>135</sup>

The previous obtained results by the QM/MM and MM MD simulations show also structural differences and H-bond changes especially in the DC gate. Hence, the computation of vibrational frequencies makes a direct comparison with the experiment possible. Furthermore, this study can also reveal spectroscopic silent substates which might additional occur between P<sub>1a</sub> and P<sub>1b</sub>, but are not visible for the experiments.

For the computation of the vibrational frequencies, DFTB with the 3ob parameter set is chosen. A previous benchmark reflects that DFTB can be successfully applied in vibrational spectroscopy studies in which H-bond systems are involved,<sup>237</sup> and on QM/MM settings.<sup>177,178</sup> To account also for an improved description of H-bonds, an extensive benchmark using DFTB 3ob i.a. on bR confirmed the good performance of DFTB reproducing qualitatively vibrational frequencies for the explanation of experimental results.<sup>228</sup> Within this study the 7 cm<sup>-1</sup> red shift observed from the dark state to the first intermediate state (1740(-)

)/1733(+) cm<sup>-1</sup>) in bR could be explained originating from distance changes between D115 and T90 as well as the different retinal configuration, while D115 and T90 are the analogue to the DC gate of ChR2-WT.

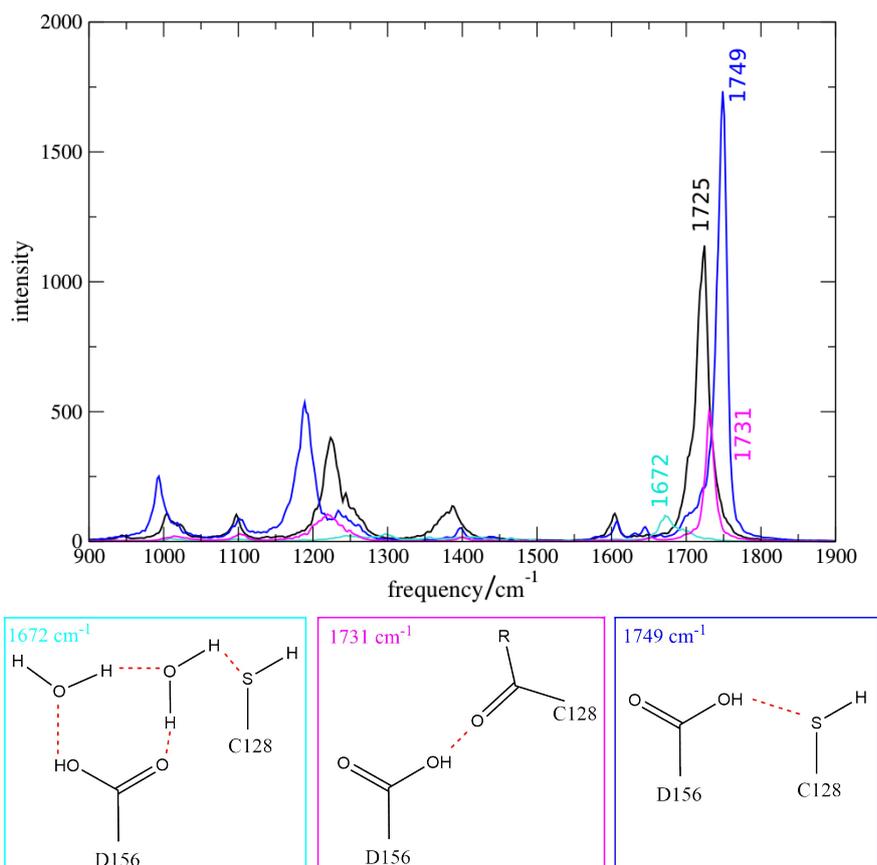
Therefore, different DC gate pattern from both QM/MM MD simulations of the early P<sub>1a</sub> and the late P<sub>1b</sub> state are chosen to calculate the vibrational frequencies by FTTCF using DFTB3/3ob, see Fig. 6.17 and Fig. 6.18.



**Figure 6.17** Computed vibrational spectra using the structural motifs of the DC gate as obtained by the QM/MM MD simulations of the P<sub>1a</sub> state. The colored border of the structural motifs corresponds to the colored band in the spectrum.

Fig. 6.17 shows the results obtained for the DC gate motifs found in the QM/MM MD simulations of P<sub>1a</sub> of the X-ray structure. The vibrational frequency of the water-bridged DC gate in the dark state of the homology model ChR2-WT was computed at 1730 cm<sup>-1</sup>.<sup>73</sup> In the first intermediate state a band at 1725 cm<sup>-1</sup> is obtained, which however shows a red shift and not the expected blue shift. The result however might be analogous to bR. Due to the stable water-bridged DC gate only minor changes like distance variation or the changed retinal configuration can effect such a shift.

The water-bridge DC gate of the X-ray structure obtains in the P<sub>1</sub> state the same band as by the homology model, i.e. at 1725 cm<sup>-1</sup>. The other DC gate motifs as shown in red and orange in Fig. 6.17 revealed however additional red shifted bands. Thus, none of these structure reveal the blue shift of about 8 cm<sup>-1</sup>, which changes was assigned to the early P<sub>1a</sub> state. However, experimental results also suggest that this band might appear after 100 ps. Hence, the used DC gate motifs here could be even to early in time.<sup>135,161</sup> Due to the steric clash with W223 and the DC gate a weakened water-bridged DC gate might induce such a blue shift.



**Figure 6.18** Computed vibrational spectra using the structural motifs of the DC gate as obtained by the QM/MM MD simulations of the P<sub>1b</sub> state. The colored border of the structural motifs corresponds to the colored band in the spectrum. The black band in the calculated spectrum corresponds to the water-bridged DC gate as shown in Fig. 6.17.

Fig. 6.18 shows the result of the DC gate motifs used for the study of P<sub>1b</sub> as obtained after 400 ns MM MD simulations of the X-ray structure. The DC gate motif containing a water cluster shows only a red shift with less intensity. The DC gate motif, where D156 is bonded to the backbone of C128 is found to give a similar band as obtained for the water-bridged DC gate in the dark state and thus account for no changes.

A large blue shift is however observed by the direct DC gate resulting from the pre-gating

process described in section 6.4. The obtained  $\nu_{COOH}$  band is about  $19\text{ cm}^{-1}$  blue shifted to the dark state and thus describes qualitatively the experimental observed blue shifted band ( $1737(-)/1760(+)\text{ cm}^{-1}$ ). In general, the frequency range of -SH bonded carboxylic acids are expected in a range of  $1745\text{--}1760\text{ cm}^{-1}$ , since the H-bonds involving sulfur are normally weaker than their oxygen counterparts.<sup>163</sup> Hence, this additional is in accordance to the observed results and the obtained direct DC gate can be assigned to the late  $P_{1b}$  state. Nevertheless, the other DC gate motifs are also involved in the previous described pre-gating process and are accompanied as spectroscopic silent substates.

## 6.8 Conclusion

The investigation of the first intermediate state  $P_1$  of ChR2-WT using QM/MM as well as MM MD simulations, reveals an insight into several structural changes as well as H-bond network changes compared to the dark state induced by the isomerized retinal upon photon absorption. The homology model as well as the X-ray structure containing more water than experimentally obtained, are used for the investigation of  $P_1$ .

The first intermediate state  $P_1$  is involved in the pre-gating process of the photocycle and divided into two substates  $P_{1a}$  and  $P_{1b}$ .<sup>135</sup> The QM/MM MD simulations of the early substate  $P_{1a}$  reveal a still weakly hydrogen bonded  $RSBH^+$ , which confirms the experimental results.<sup>135</sup> Two different H-bond pattern are found of the  $RSBH^+$ : it forms a H-bond either to (i) T127 ( $RSBH^+ \cdots O\text{-(T127)}$ ) or (ii) S256 ( $RSBH^+ \cdots O\text{-(S256)}$ ), dependent whether described by QM/MM or MM methods. The first one is obtained in QM/MM, while the second one is found in MM MD simulations. With this investigation, the choice of the 13-*cis* retinal configuration for the study of  $P_1$  becomes also important. In this study, we used a 13-*cis* retinal configuration, originated from a recent CASPT2//CASSCF/MM study on the retinal isomerization pathway of ChR2-WT.<sup>74</sup> Previous computational studies neglected excited state effects on the retinal configuration when modeling the isomerized retinal. Even though, Cheng *et al.*<sup>37</sup> obtained the same H-bond pattern of the  $RSBH^+$ , namely  $RSBH^+ \cdots O\text{-(T127)}$ , since the used dihedral angle is similar to our used 13-*cis* retinal. However, Ardevol *et al.*<sup>13</sup> showed another  $RSBH^+$  H-bond pattern, namely ( $RSBH^+ \cdots O\text{-(D253)}$ ), using MM MD simulations. This H-bond pattern is also found when performing MM MD simulations on a less equilibrated homology model system. Hence, this might be either a forcefield artefact or originates from too less equilibration.

Additionally, the H-bond pattern  $RSBH^+ \cdots O\text{-(D253)}$  has also influences on the further processes. While in QM/MM as well as MM MD simulations a steric clash by the 13-*cis* retinal and W223 is found, the simulation containing the H-bond pattern  $RSBH^+ \cdots O\text{-(D253)}$  lead to a delayed steric clash, dependent how long the  $RSBH^+$  is hydrogen bonded to D253. Ardevol *et al.*<sup>13</sup> reported also this steric clash, but assigned it to the early  $P_2$  state. However, this would be in conflict with the experimental results, since the formation of the

band obtained by structural rearrangements in the binding pocket of the isomerized retinal, appeared directly after retinal isomerization.<sup>134, 135, 161, 183</sup> Thus, our results agree with the experimental observation as well as to previous computational studies,<sup>208</sup> reported also a steric clash between the 13-cis retinal and W223.

This steric clash is accompanied with the displacement of the indol ring of W223 from the C<sub>13</sub>- to the C<sub>9</sub>-methyl group of retinal, which leads in turn to a steric clash with the DC gate, followed by the formation of a direct hydrogen bonded DC gate. These structural rearrangements lead to the subsequent formation of a cavity between the DC gate and retinal, making water influx possible. In the case of the homology model more water is influxed, so that C128 is pushed outward, which in turn lead to a slightly kink of helix C.

Hence, the MM MD simulations sampled several possible substates differing mainly in the DC gate motifs. These structures are further used for QM/MM MD simulations referred as to P<sub>1b</sub>. The computation of the UV/Vis spectra of the sampled structure by QM/MM MD simulations of both P<sub>1a</sub> and P<sub>1b</sub> reveals the same absorption maxima range for both. Furthermore, a red-shift is obtained compared to the dark state, which is qualitatively in agreement with the experiment.<sup>134</sup> Analogous to the dark state of ChR2-WT, a high flexibility of the structure is found for the P<sub>1</sub> state leading to a broad absorption maxima range.

The sampled substates differing in the DC gate motifs are further used for the computation of vibration spectra using FTTCF and DFTB3/3ob. Hence, P<sub>1b</sub> could be identified. The computed vibrational band of the direct hydrogen bonded DC gate account for the high blue shift from the dark state to P<sub>1</sub>, which was experimentally observed and assigned to the late substate P<sub>1b</sub>.<sup>135</sup>

The comparison of the results obtained by the homology model and the X-ray structure shows in several structural motifs similarities as well as in H-bond network changes. One main difference is obtained due to the orientation of E90, which is in the X-ray structure downward oriented towards the extracellular side. QM/MM MD simulations found no H-bond changes in the homology model regarding the central gate. However, in the X-ray structure H-bond changes are observed. While in the dark state E90 is mainly hydrogen bonded to water molecules, bridging the counterions E123 and D253 with E90, in P<sub>1</sub> and especially in P<sub>1b</sub> a direct hydrogen bond of E90 to E123 or D253 is found. This is also in accordance to experimental results, suggesting a blue shift of the carboxylic band compared to the dark state, which is assigned to E90.<sup>133, 134</sup> No significant influence of the central gate with the downward oriented E90 is obtained by MM MD simulation with respect to the reported pre-gating mechanism. In contrast, in several MM MD simulations of the homology model a downward flip of E90 is found. This is accompanied with the formation of a water pore between helices A, B and G, which was already reported by previous computational studies,<sup>13, 37, 120</sup> assigning the downward flip of E90 to H-bond changes in the central gate. However, this could be only the case in C1C2 or the homology model based on C1C2, but is not a possible process in the ChR2-WT, since E90 is already downward oriented in the dark state. This is in accordance

to experimental results suggesting differences in the channel mechanism and functionality of C1C2 and ChR2-WT.<sup>87</sup>

In contrast, the MM MD simulations of the homology model containing no flip of E90 and the X-ray structure, reveal a water pore formation between helices A-B-C-D-G. However, this water pore is only formed on one half part of the channel, i.e. from the extracellular side up to retinal. The formation or enlargement of water-filled pores inside the protein is indeed valid, since experimental studies suggested it to be a prerequisite for cation permeation.<sup>132, 134</sup>

Furthermore, this water influx lead to a structural motif, which contains a water bridge between the RSBH<sup>+</sup> and E123. Additionally, E123 is hydrogen bond bridged via K93 to D253. This might indicate a possible deprotonation pathway of the RSBH<sup>+</sup>, i.e. a proton transfer via a water molecule to E123 and via K93 to D253, which is suggested to be the proton acceptor of the RSBH<sup>+</sup> as obtained by FTIR measurements.<sup>134</sup> Further investigation of this possible proton transfer have to be performed to see whether this can be confirmed.

In general, the investigation of the P<sub>1</sub> state shows, that a QM/MM description is necessary for a good description of hydrogen bonds and important structural motifs. The so sampled structural motifs were able to confirm the experimental results with the help of UV/Vis and vibration spectra computation, but much more they were able to give an insight into the exact structural changes, which cannot be figured out experimentally. Thus, the root of the experimentally obtained shifts could be explained and the late P<sub>1b</sub> state could be identified. These analysis in turn confirmed the suggested pre-gating mechanism. Furthermore, the homology model obtained similar results to the X-ray structure and is thus still valid for further investigations. With these obtained structures further processes like the deprotonation of the RSBH<sup>+</sup> can be investigated to model the early P<sub>2</sub> state or further intermediate states of the photocycle to understand mechanisms like the opening and closing of the channel.

---

## Conclusion and Outlook

This thesis included two main topics. One topic was the benchmark of the recent time-dependent extension of the LC-DFTB method<sup>115</sup> for its application to biological systems, specifically light-activated proteins such as rhodopsins and light-harvesting (LH) complexes. The second topic was the investigation of the pre-gating process of the light-activated ion channel Channelrhodopsin-2 wild-type (ChR2-WT) using established methods.

In chapter 4, the benchmark of LC-DFTB was first performed on the chromophores retinal and bacteriochlorophylla (BChla) focusing on different influencing factors responsible for their characteristic absorption spectra. These influencing factors included the geometry of the chromophore, the electrostatic environment of the protein, as well as exciton couplings in the case of BChla. Furthermore, several methods were used ranging from *ab-initio* to semiempiric levels of theory. The results showed, that LC-DFTB could reproduce qualitatively the same trend as the LC-DFT functionals with respect to the different influencing factors. For retinal, LC-DFT functionals as well as LC-DFTB obtained an improved description concerning the geometrical impact on excitation energies as well as the influence of the electrostatic environment compared to GGA or hybrid DFT functionals. However, they still showed a color-weakness when including the electrostatic environment compared to the reference method SORCI with respect to describe the correct magnitude of the shift. In the case of BChla, LC-DFT functionals and LC-DFTB displayed the same trend as the reference method ADC(2) regarding the geometrical impact on the excitation energies of BChla. The inclusion of the electrostatic environment reveals, that LC-DFTB could reproduce the same trend as DFT/MRCI and even better than CAM-B3LYP. In contrast to ZINDO/S, it did not overestimate the excitation energies when including the electrostatic environment. The benchmark study of the exciton coupling on an artificial system, i.e. a truncated BChla structure optimized in vacuum, showed that the Coulomb couplings obtained by LC-DFTB are slightly overestimated compared to the supermolecule calculations. The application to the biological system, i.e. the B850 ring system of the LH2 complex, revealed however that LC-DFTB could nicely reproduce qualitatively the same trend as the supermolecule

---

calculations using LC-DFT.

In chapter 5, the benchmark of LC-DFTB was then extended to its performance in rhodopsins and light-harvesting complexes. For the application to rhodopsins, bacteriorhodopsin (bR) and pharaonis phoborhodopsin (ppR) served as models. The application of LC-DFTB to both bR and ppR, revealed the same weaknesses to describe the correct magnitude of the absorption maxima shift between both proteins. This was in agreement to the results obtained by the benchmark study on retinal. In the case of the LH complexes, the light-harvesting complex II (LH2) served as a model system. Analogous to the results of the benchmark study, LC-DFTB could describe electrostatic influences in the right order showing also that the main shift originates by considering the Coulomb couplings. Furthermore, QM/MM MD simulations were performed on the LH2 complex, giving new insights in the BChl a geometry. With the QM/MM sampled geometries a too broad absorption range was obtained which does not match the experimental one. The root of this discrepancy must be investigated in a future work.

The benchmark study as a whole showed, that LC-DFTB is not suitable for the study of rhodopsins. Further developments are necessary to improve the weaknesses. In the case of LH complexes, LC-DFTB is a promising method. Due to the low computational cost, it is useful for the study of very large systems. With further benchmarks concerning the exciton couplings, LC-DFTB could be useful for further studies of the exciton transfer dynamics towards the reaction center of LH complexes.

Finally, in chapter 6 several methods, MM as well as QM/MM methods were applied to investigate the first intermediate state  $P_1$  of ChR2-WT, which is involved in the pre-gating process of the photocycle. The QM/MM MD simulations as well as MM MD simulations sampled several substates of  $P_1$ . These structures were confirmed by subsequent computation of UV/Vis spectra obtaining a red shift with respect to the dark state, which was qualitatively in agreement to the experimentally obtained shift. Furthermore, the computation of vibrational frequencies of these structures identified the  $P_{1b}$  state. This was reflected in hydrogen bond changes of the DC gate, which was in accordance to the experimentally observed weakening of the hydrogen bond in the DC gate. Furthermore, from the MM MD simulations a pre-gating mechanism is revealed, which indicated the formation of a water pore. With this study, not only the experimental results are confirmed, but figured out several structural motifs and hydrogen bond network changes, which cannot be determined experimentally. Hence, these results are helpful for further investigations of other intermediate states of the photocycle as well as the understanding of the channel mechanism and functionality.

---

## Appendix

### A.1 Benchmark of LC-DFTB on Retinal and Bacteriochlorophyll

#### A.1.1 Retinal

**Table A.1** Excitation energies (in eV) of all-*trans* retinal geometries optimized in vacuum using different methods. Bond length alternations are given in Å.

	CASSCF	HF	BH-LYP	B3LYP	DFTB	BLYP
Bond length alternation <sup>a</sup>	0.100	0.069	0.035	0.028	0.025	0.023
SORCI <sup>a</sup>	2.10	2.07	2.04	1.95	1.91	1.89
ZINDO/S	2.36	2.38	2.33	2.28	2.25	2.24
CAM-B3LYP	2.68	2.77	2.80	2.76	2.69	2.68
LC-BLYP	2.75	2.76	2.70	2.62	2.57	2.56
$\omega$ B97X	2.82	2.82	2.76	2.67	2.62	2.61
LC-DFTB	2.63	2.69	2.69	2.60	2.57	2.55
ADC(2)	1.94	2.01	2.03	1.93	1.89	1.87
SOS-ADC(2)	2.32	2.28	2.20	2.08	2.05	2.01

<sup>a</sup> Taken from Ref. 222.

**Twist of  $\beta$ -ionone ring****Table A.2** Excitation energies (in eV) of a 6-s-cis-11-cis PSB in vacuum for different twist angle of the  $\beta$ -ionone ring.

method	33°	60°	90°	171°
SORCI <sup>a</sup>	1.96	1.99	2.28	1.95
ZINDO/S	2.32	2.40	2.58	2.27
CAM-B3LYP	2.68	2.66	3.08	2.70
LC-BLYP	2.63	2.70	2.94	2.60
WB97X	2.69	2.77	3.00	2.66
LC-DFTB	2.59	2.57	2.94	2.59
ADC(2)	1.94	2.03	2.27	1.92
SOS-ADC(2)	2.15	2.26	2.40	2.08

<sup>a</sup> Taken from Ref. 222.**A.1.2 Bacteriochlorophyll****Previous Computational Studies - Exciton Couplings**

Two tables are given in the following, concerning previous studies of supermolecule calculations and Coulomb couplings on the LH2 complex. The first table shows the exciton couplings on a static model, while the second table summarizes the studies based on MD simulations.

**Table A.3** Previous studies on exciton couplings on the X-ray structure of the LH2 complex using different methods. The Coulomb couplings of the strongest BChl a dimers are displayed.

Reference	X-ray structure	Methods		Coulomb couplings	
		excitation energy	exciton coupling	cm <sup>-1</sup>	eV
Krüger et al., <sup>117</sup> (1998)	LH2 <i>Rps. ac.</i>	CIS (3-21G*)	TDC	238	0.03
Cory et al., <sup>39</sup> (1998)	LH2 <i>Rs. ms.</i>	INDO-CIS	INDO-CIS effective Hamiltonian	790	0.10
Scholes et al., <sup>190</sup> (1999) <sup>a</sup>	LH2 <i>Rps. ac.</i>	CIS	CIS (6-31G*) (scaled exp.)	320	0.04
Madjet et al., <sup>139</sup> (2006)	LH2 <i>Rps. ac.</i>	CIS, TD-DFT/B3LYP (6-31G*)	TrESP, CIS TrESP,TD-DFT/B3LYP	245	0.03
Neugebauer, <sup>159</sup> (2008)	LH2 <i>Rps. ac.</i>	TD-DFT +FDE (SOAP/TZP)	supermol. TD-DFT + FDE supermol. TD-DFT <sup>b</sup> FDE <sup>c</sup>	242	0.03
Cupellini et al., <sup>42</sup> (2016)	LH2 <i>Rps. ac.</i>	TD-DFT + MMPOL, CAM-B3LYP (6-31G(d))	TD-DFT + MMPOL by EXAT <sup>c</sup>	409	0.05
Segatta et al., <sup>192</sup> (2017)	LH2 <i>Rps. ac.</i>	RASSCF/RASPT2	TDC <sup>c</sup>	563	0.07
		TD-DFT + MMPol CAM-B3LYP	TD-DFT + MMPol	474	0.06
			TD-DFT + MMPol	336	0.04
				288	0.04

<sup>a</sup>also performed supermol. calc. CIS(6-31G\*); 0.09 eV (730 cm<sup>-1</sup>) and 0.07 eV (550 cm<sup>-1</sup>); <sup>b</sup>DFT  $E_{xc}$  functional of GGA type: exchange by Becke, and correlation functional by PW91;<sup>159</sup> <sup>c</sup>EXAT is the program used for computing the excitonic Hamiltonian, cf. Ref. 42,96; <sup>d</sup>no explicit screening factor used

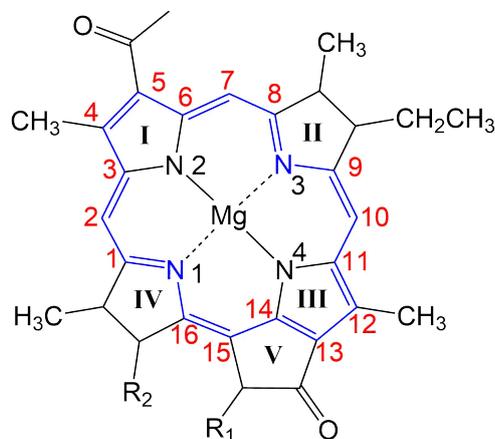
**Table A.4** Previous studies on Coulomb couplings of the LH2 complex using different methods. The Coulomb couplings are computed on sampled structures by classical MD simulations. Values of the strongest BChl a dimers are displayed.

Reference	X-ray structure and MD	Methods		Coulomb couplings	
		excitation energy	Coulomb coupling	cm <sup>-1</sup>	eV
Olbrich et al., <sup>171</sup> (2011)	LH2 <i>R<sub>s</sub></i> , <i>m<sub>s</sub></i> , 12 ps	ZINDO/S-CIS	PDA <sup>a</sup>	274	0.03
			T-ESP, TD-DFT/B3LYP	173	0.02
				140	0.02
Cupellini et al., <sup>42</sup> (2016)	LH2 <i>R<sub>ps</sub></i> , <i>a<sub>c</sub></i> , 100ns	TD-DFT + MMPol, CAM-B3LYP (6-31G(d))	TD-DFT + MMPol by EXAT <sup>b</sup>	339	0.04
				317	0.04

<sup>a</sup> transition densities obtained by ZINDO/S-CIS

<sup>b</sup> EXAT is the program used for computing the excitonic Hamiltonian, cf. Ref.<sup>42,96</sup>

## Bond Length Alternation



**Figure A.1** In blue diaza[18]-annulene substructure of BChl a;  $R_1 = \text{H}$  (truncated), COOMe;  $R_2 = \text{CH}_3$  (truncated), COOMe (Phytyl-Tail truncated); Phytyl. The C atom numbers are displayed in red, the N atom numbers in black.

**Table A.5** Excitation energies (in eV) of BChl a in vacuum. D1 values are given in parenthesis.

	HF	CAM-B3LYP	BH-LYP	B3LYP	DFTB	BLYP
BLA (Å)	0.102	0.032	0.019	0.005	0.004	0.004
TD-BP86	2.06	2.09	2.15	2.09	2.06	2.04
TD-DFTB	1.82	1.81	1.86	1.81	1.80	1.77
TD-B3LYP	2.18	2.12	2.17	2.14	2.12	2.10
HF/CIS	3.33	2.36	2.12	1.87	1.84	1.81
ZINDO/S	2.14	1.57	1.51	1.47	1.46	1.45
CAM-B3LYP	2.47	2.03	2.04	1.99	1.97	1.95
LC-BLYP	2.78	2.03	1.95	1.86	1.83	1.82
$\omega$ B97X	2.83	2.05	1.97	1.87	1.84	1.83
LC-DFTB	2.40	1.91	1.93	1.85	1.83	1.80
LC-DFTB*	2.37	1.91	1.93	1.87	1.85	1.81
SOS-ADC(2)*	2.40 (0.0577)	1.44 (0.0834)	1.54 (0.0796)	1.75 (0.0562)	1.73 (0.056)	1.67 (0.0639)
SOS-CC2*	2.48 (0.1204)	1.68 (0.1959)	1.75 (0.1854)	1.84 (0.0960)	1.83 (0.0968)	1.81 (0.1150)

\* Modified side chains.

**Table A.6** Bond lengths of BChl a substructure, shown in A.1 in blue.

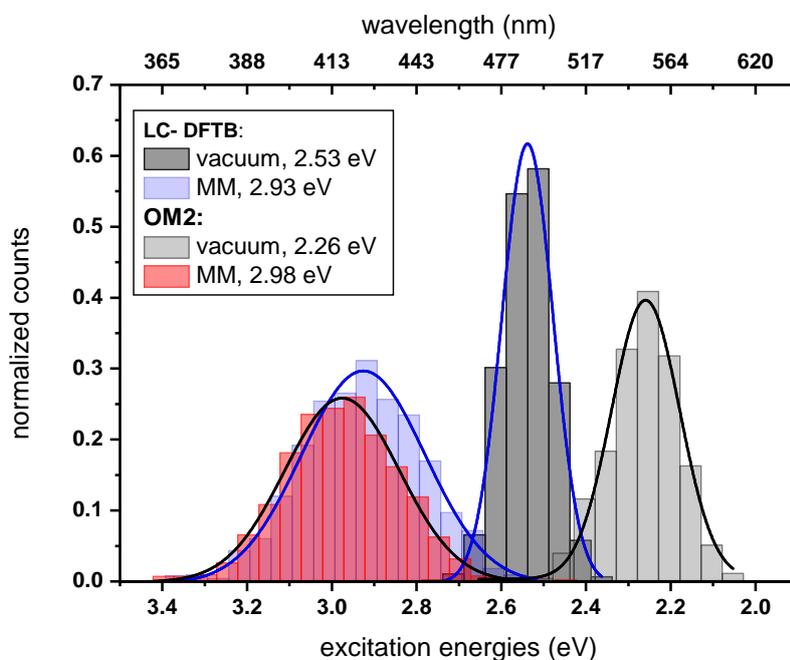
	HF	CAM-B3LYP	BH-LYP	B3LYP	DFTB	BLYP
C1=N1:	1.293	1.340	1.337	1.358	1.364	1.372
C1-C2:	1.432	1.399	1.388	1.391	1.392	1.401
C2=C3:	1.366	1.397	1.397	1.413	1.416	1.423
C3-C4:	1.488	1.456	1.444	1.446	1.440	1.454
C4=C5:	1.342	1.377	1.380	1.401	1.398	1.415
C5-C6:	1.489	1.465	1.453	1.459	1.454	1.470
C6=C7:	1.359	1.393	1.395	1.412	1.413	1.422
C7-C8:	1.439	1.403	1.390	1.392	1.393	1.401
C8=N3:	1.289	1.335	1.335	1.356	1.363	1.370
N3-C9:	1.412	1.376	1.360	1.365	1.370	1.377
C9=C10:	1.339	1.373	1.375	1.392	1.393	1.402
C10-C11:	1.455	1.418	1.408	1.412	1.417	1.421
C11=C12:	1.389	1.431	1.433	1.451	1.446	1.461
C12-C13:	1.426	1.398	1.387	1.395	1.399	1.407
C13=C14:	1.377	1.407	1.405	1.420	1.421	1.431
C14-C15:	1.449	1.416	1.404	1.408	1.411	1.418
C15=C16:	1.335	1.369	1.372	1.389	1.396	1.400
C16-N1:	1.414	1.382	1.363	1.368	1.373	1.381
<b>BLA</b> (Å):	0.102	0.032	0.019	0.005	0.004	0.004

**Table A.7** Dihedral angles of BChl a substructure, shown in A.1 in blue.

Dihedral angle	HF	CAMB3LYP	BHLYP	B3LYP	DFTB	BLYP
N1-C1-C2=C3:	3.28	7.90	1.92	2.01	0.76	2.18
N2-C6-C7=C8:	0.28	-2.02	-0.55	-0.76	1.86	-0.85
N3-C9-C10=C11:	-2.17	4.92	-1.71	-1.89	-1.51	-1.94
N4-C14-C15=C16:	1.79	-0.28	1.72	1.70	2.30	1.78

## A.2 Performance of LC-DFTB on Rhodopsins and Light-Harvesting Complex II

### A.2.1 Rhodopsins



**Figure A.2** Simulated absorption spectrum of ppR. LC-DFTB and OM2/MRCI are used for the computation of the excitation energies. The histograms are based on snapshot geometries of QM/MM MD trajectories (1 ns length). Plotted are the excitation energies weighed by the oscillator strength for (i) only the retinal chromophore (vacuum) and (ii) with additional fixed MM point charges to account for the protein environment (MM). Gaussian functions are used to determine the corresponding maxima, in blue: LC-DFTB and in black: OM2/MRCI.

## A.2.2 Light-Harvesting Complex II

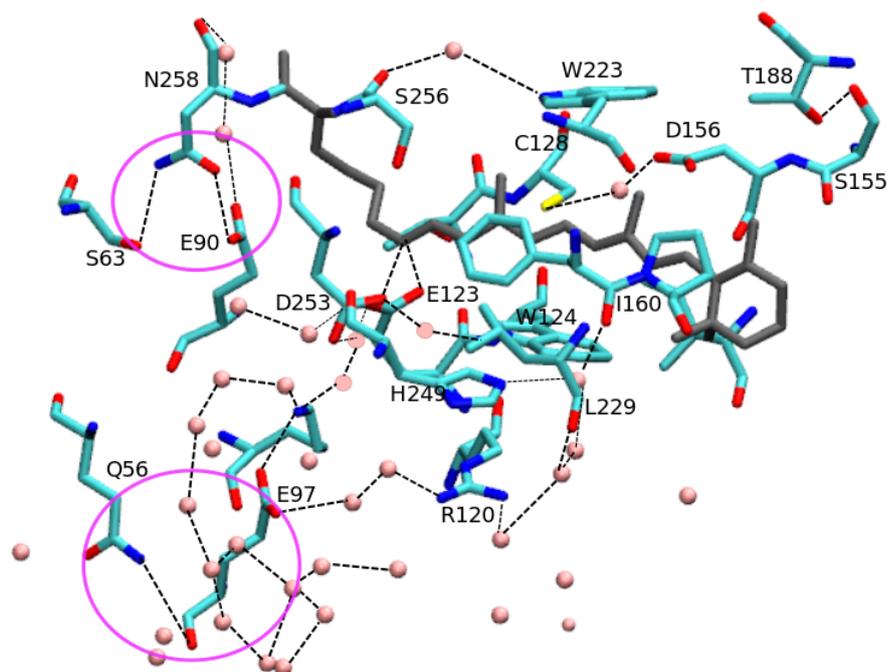
## QM/MM optimized model

**Table A.8** BLA of the QM/MM optimized BChla chromophores of the LH2 complex. Excitation energies (in eV) calculated in vacuum and the resulted shift (protein-vacuum) using LC-DFTB and ZINDO/S.

	BLA (Å)	LC-DFTB		ZINDO/S	
		vacuum	shift (eV)	vacuum	shift (eV)
B800					
1	0.010	1.83	+0.01	1.46	+0.06
2	0.013	1.84	+0.05	1.46	+0.14
3	0.006	1.82	+0.03	1.46	+0.12
4	0.007	1.83	-0.02	1.46	-0.02
5	0.014	1.84	±0.00	1.47	-0.01
6	0.010	1.82	-0.03	1.45	-0.03
7	0.011	1.82	+0.01	1.46	+0.03
8	0.011	1.83	±0.00	1.47	+0.02
B850					
1	0.005	1.83	+0.02	1.46	+0.04
2	0.001	1.83	±0.00	1.45	±0.00
3	0.008	1.83	+0.01	1.46	+0.01
4	0.003	1.83	-0.01	1.45	+0.01
5	0.007	1.83	+0.01	1.46	+0.02
6	0.004	1.83	±0.00	1.46	+0.01
7	0.003	1.83	±0.00	1.46	±0.00
8	0.003	1.82	-0.01	1.45	-0.00
9	0.005	1.83	-0.01	1.46	+0.01
10	0.005	1.83	+0.01	1.45	+0.01
11	0.005	1.83	+0.02	1.46	+0.06
12	0.002	1.83	-0.01	1.45	-0.00
13	0.005	1.83	-0.01	1.46	-0.01
14	0.004	1.83	-0.01	1.45	-0.01
15	0.005	1.83	±0.00	1.46	+0.00
16	0.002	1.82	±0.00	1.45	+0.02

## A.3 The First Intermediate State of Channelrhodopsin-2 Wild-Type

### Hydrogen Bond Network of the Binding Pocket of Retinal: Homology Model versus X-Ray Structure



**Figure A.3** Binding pocket of retinal: homology model versus X-ray

#### A.3.1 QM/MM MD Simulations of the Early Substate $P_{1a}$

##### DC Gate

In the subsequent table, the percentage of the H-bond pattern occurring in the DC gate as observed by the QM/MM MD simulations of the early substate  $P_{1a}$  are presented. Additionally, the distances between the center of mass (COM) of the  $C_{13}$ -methyl group of retinal and the COM of the indol ring of W223 are measured during all QM/MM MD simulations.

**Table A.9** Percentage of the H-bond occurring in the DC gate as observed by QM/MM MD simulations of the early substrate P<sub>1a</sub>. And distance between the COM of the C<sub>13</sub>-methyl group of retinal and the COM of the indol ring of W223.

Model	structural motif of the dark state	QM region (D156)O...H <sub>2</sub> O	D156...H <sub>2</sub> O...C128	distance W223 <sup>d)</sup>	
homology model	dark state	QM-T <sup>a)</sup>	98.8 <sup>b)</sup>	32.6 <sup>b)</sup>	3.85
	system HW <sup>c)</sup>	QM	-	-	3.29
		QM-TS	-	-	3.33
	system LW <sup>c)</sup>	QM	99.1	28.1	3.33
		QM-TS	98.6	33.8	3.30
		QM-DC	95.2	34.3	3.57
		QM-E90	94.8	34.0	3.60
X-ray	E123 <sup>down</sup> -RSBH <sup>+</sup> ...-O(D253)	QM-E90	97.7	34.1	3.20
	E123 <sup>down</sup> -RSBH <sup>+</sup> ...H <sub>2</sub> O		97.5	33.2	3.23
	E123 <sup>up</sup> -RSBH <sup>+</sup> ...H <sub>2</sub> O		92.5	26.0	-
	E123 <sup>up</sup> -RSBH <sup>+</sup> ...-O(E123)		95.8	34.7	3.31

<sup>a)</sup> QM-T is the QM region used in see Ref. 73; <sup>b)</sup> QM/MM MD trajectories used in Ref. 73, values calculated by F. Wolff;

<sup>c)</sup> E123<sup>up</sup>-RSBH<sup>+</sup>...-O(E123); <sup>d)</sup> only values are taken of the average for QM/MM trajectory with no turn of W223 towards C<sub>9</sub>-methyl group of retinal.

## Central Gate

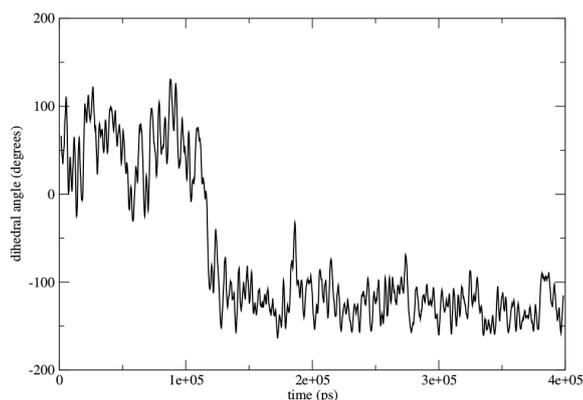
**Table A.10** Percentage of the H-bond pattern (%) of the central gate of the homology model.

Model	structural motif of the dark state	QM region	E90-H <sub>2</sub> O	E90-N258	N258-S63
homology model	system HW <sup>a)</sup>	QM	89.0	99.7	74.5
		QM-TS	85.0	99.8	80.3
	system LW <sup>a)</sup>	QM	88.9	99.6	56.0
		QM-TS	42.4	99.7	54.5
		QM-DC	45.0	99.6	54.2
		QM-E90	40.8	99.3	89.2

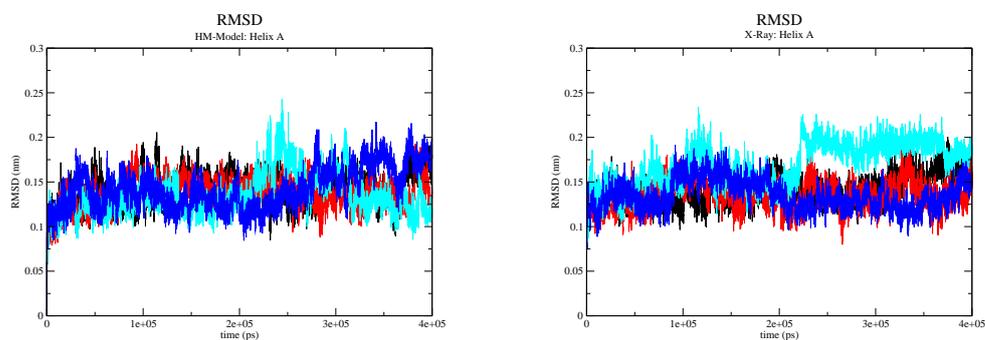
<sup>a)</sup>E123<sup>up</sup>-RSBH<sup>+</sup> . . . -O(E123)

A.3.2 MM MD Simulations of P<sub>1</sub>

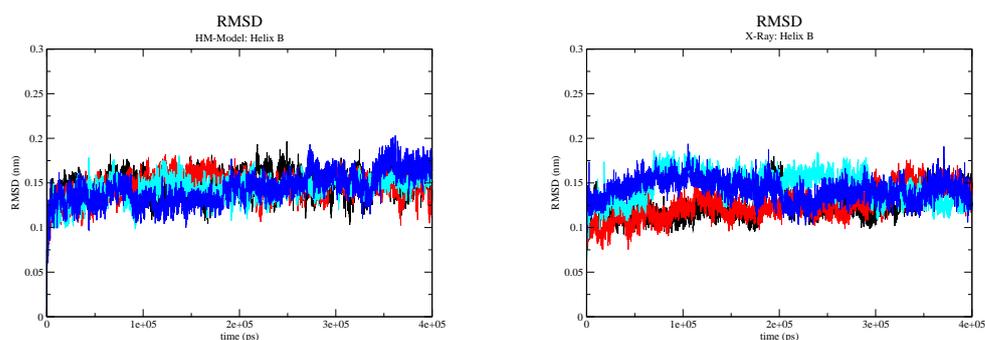
## Pre-Gating Process

**Figure A.4** Dihedral angle C<sub>10</sub>-C<sub>11</sub>-C<sub>12</sub>-C<sub>13</sub> of 13-*cis* retinal measured over one MM MD simulation of system LW, where the turn of the C<sub>13</sub>-methyl group is found.

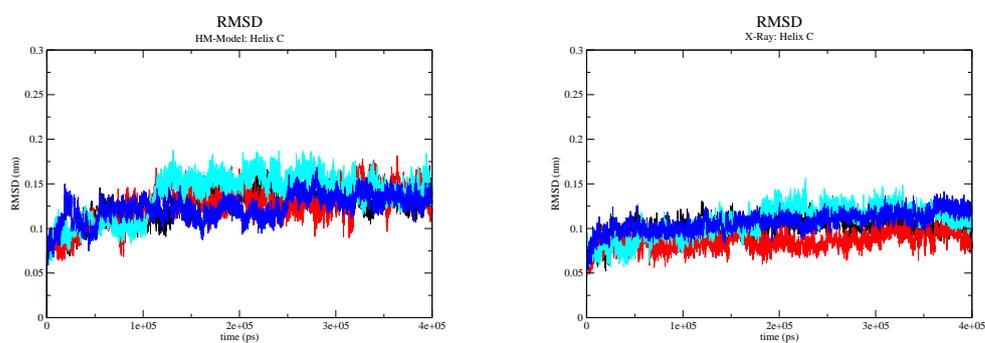
## RMSD



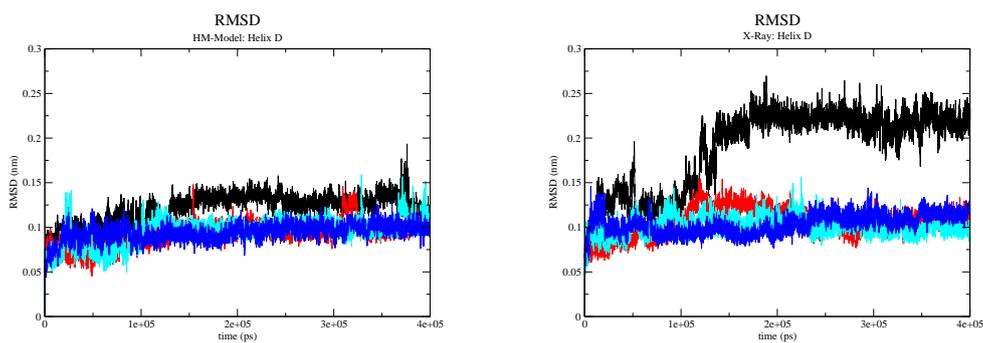
**Figure A.5** RMSD of helix A; Left: homology model, system LW and Right: X-ray structure; Color code: black: trajectory 1, red: trajectory 2, cyan: trajectory 3 and blue: trajectory 4.



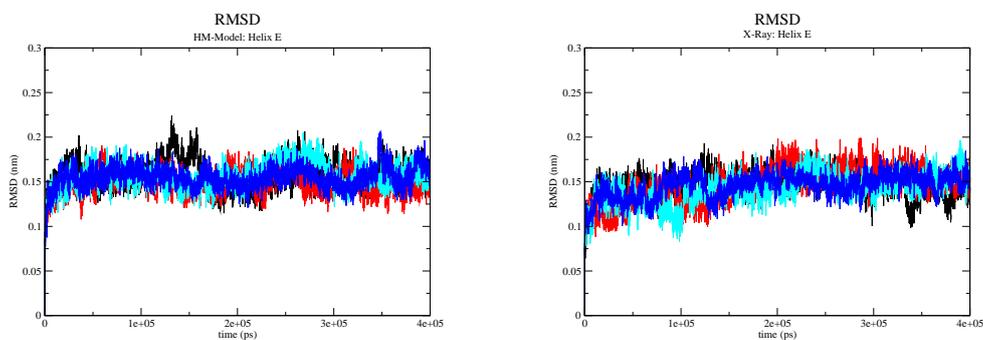
**Figure A.6** RMSD of helix B; Left: homology model, system LW and Right: X-ray structure; Color code: see Fig. A.5.



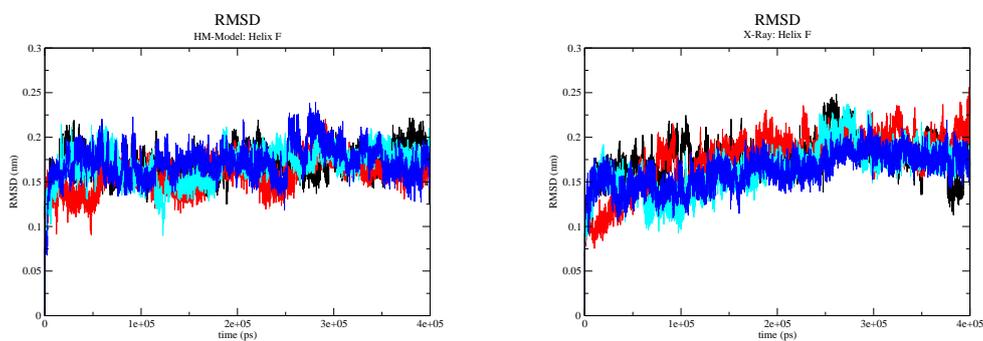
**Figure A.7** RMSD of helix C; Left: homology model, system LW and Right: X-ray structure; Color code: see Fig. A.5.



**Figure A.8** RMSD of helix D; Left: homology model, system LW and Right: X-ray structure; Color code: see Fig. A.5.



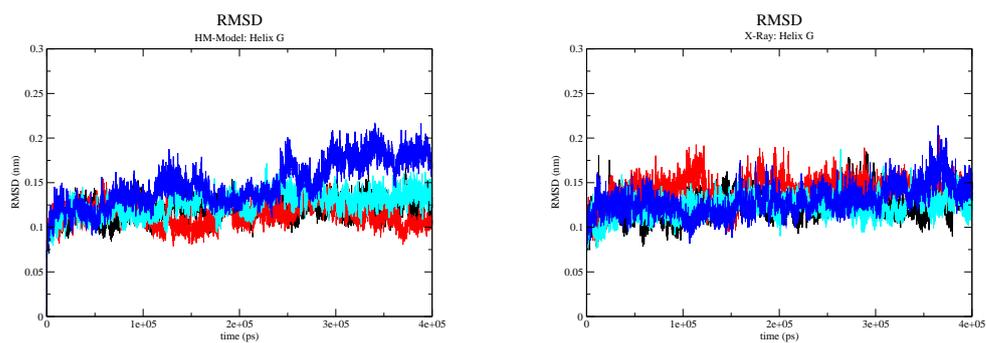
**Figure A.9** RMSD of helix E; Left: homology model, system LW and Right: X-ray structure; Color code: see Fig. A.5.



**Figure A.10** RMSD of helix F; Left: homology model, system LW and Right: X-ray structure; Color code: see Fig. A.5.

### A.3. The First Intermediate State of Channelrhodopsin-2 Wild-Type

---



**Figure A.11** RMSD of helix G; Left: homology model, system LW and Right: X-ray structure; Color code: see Fig. A.5.

---

## List of Abbreviations

<b>AO</b>	Atomic Orbital
<b>BChl</b>	Bacteriochlorophyll
<b>BLA</b>	Bond Length Alternation
<b>bR</b>	Bacteriorhodopsin
<b>ChR</b>	Channelrhodopsin
<b>CI</b>	Configuration Interaction
<b>CAS</b>	Complete Activ Space
<b>CHARMM</b>	Chemistry at Harvard Macromolecular Mechanics
<b>DFT</b>	Density Functional Theory
<b>DFTB</b>	Density-Functional Tight-Binding
<b>FT-IR</b>	Fourier-Transformed Infrared
<b>FSD</b>	Fourier Self-Deconvolution
<b>GGA</b>	Generalized Gradient Approximation
<b>GROMACS</b>	Groningen Machine for Chemical Simulation
<b>HF</b>	Hartree Fock
<b>HK</b>	Hohenberg-Kohn
<b>HOMO</b>	Highest Occupied Molecular Orbital
<b>IR</b>	Infrared
<b>LUMO</b>	Lowest Unoccupied Molecular Orbital
<b>KS</b>	Kohn-Sham
<b>LCAO</b>	Linear Combination of Atomic Orbitals
<b>LDA</b>	Local Density Approximation
<b>LH</b>	Light-Harvesting
<b>MD</b>	Molecular Dynamic
<b>MO</b>	Molecular Orbital
<b>MM</b>	Molecular Mechanics
<b>MP</b>	Møller and Plesset
<b>MRCI</b>	Multi-Reference Configuration Interaction
<b>NDDO</b>	Neglect of Diatomic Differential Overlap
<b>OM2</b>	Orthogonalization Model 2
<b>QM</b>	Quantum Mechanics
<b>QCL</b>	Quantum Cascade Laser
<b>RAS</b>	Restricted Active Space
<b>RHF</b>	Restricted Hartree Fock
<b>Rho</b>	Rhodopsin
<b>SB</b>	Schiff Base
<b>SORCI</b>	Spectroscopy Oriented Configuration Interaction

<b>SCF</b>	Self-Consistent Field
<b>UV</b>	Ultraviolet
<b>VdW</b>	Van der Waal
<b>VMD</b>	Visual Molecular Dynamics
<b>Vis</b>	Visible
<b>WT</b>	Wild-Type
<b>ZDO</b>	Zero Differential Overlap
<b>eV</b>	Electronvolt
<b>fs</b>	Femtosecond
<b>nm</b>	Nanometer
<b>ns</b>	Nanosecond
<b>ps</b>	Picosecond
<b>Å</b>	Ångström

---

## List of Figures

2.1	Spectrum of Electromagnetic Radiation . . . . .	3
2.2	Franck-Condon Principle . . . . .	4
2.3	Schematic Illustration of Photochemical and Photophysical Processes . . . . .	6
2.4	Possible Exciton Transfer Mechanisms . . . . .	8
2.5	Retinal and $\beta$ -Carotene . . . . .	9
2.6	Bacteriochlorophyll a and Porphyrin . . . . .	10
2.7	Rhodopsin Embedded in Membrane and Isomerization of Retinal . . . . .	11
2.8	Photocycle of bR and Active Site of bR and ppR . . . . .	12
2.9	ChR2-WT Structural Motifs . . . . .	14
2.10	ChR2-WT Photocycle . . . . .	15
2.11	FMO Trimer and FMO Monomer . . . . .	17
2.12	LH1 Complex and LH2 Complexes . . . . .	18
2.13	LH2 Complex Structural Motifs . . . . .	20
3.1	Timescales of Biological Processes . . . . .	21
3.2	CAS and RAS . . . . .	26
3.3	Excitonic Splitting . . . . .	47
4.1	All- <i>trans</i> and 11- <i>cis</i> Retinal . . . . .	53
4.2	Bacteriochlorophyll a . . . . .	55
4.3	Retinal: Excitation Energies vs. BLA . . . . .	62
4.4	Retinal: Twist of $\beta$ -Ionone Ring . . . . .	64
4.5	PES of 6- <i>s-cis</i> -to 6- <i>s-trans</i> -PSB11 . . . . .	65
4.6	PSB5 . . . . .	65
4.7	BLA of BChl a in Vacuum . . . . .	67
4.8	BChl a: Excitation Energies vs. BLA in Vacuum . . . . .	68
4.9	Two BChl a Configurations of FMO . . . . .	70
4.10	BChl a: Environmental Effects on Excitation Energies . . . . .	70
4.11	BChl a: Environmental Effects on Excitation Energies, Shifts . . . . .	71
4.12	Truncated BChl a Dimers . . . . .	72
4.13	BChl a Dimer Type of the B850 Ring of the LH2 Complex . . . . .	76

---

5.1	Simulated Absorption Spectra of bR . . . . .	88
5.2	BLA of one BChl a of the B800 Ring, MM and QM/MM MD . . . . .	92
5.3	LC-DFTB: BLA vs. Excitation Energy . . . . .	93
5.4	ZINDO/S: BLA vs. Excitation Energy . . . . .	93
5.5	LC-DFTB: Environmental Effects on the Excitation Energies of MM and QM/MM sampled BChl a Structures (B800) . . . . .	94
5.6	LC-DFTB: Environmental Effects on the Excitation Energies of MM and QM/MM sampled BChl a Structures (B850) . . . . .	94
5.7	ZINDO/S: Environmental Effects on the Excitation Energies of MM and QM/MM sampled BChl a Structures (B800) . . . . .	95
5.8	ZINDO/S: Environmental Effects on the Excitation Energies of MM and QM/MM sampled BChl a Structures (B850) . . . . .	95
5.9	MM MD simulation and LC-DFTB: Absorption Spectra of B800 und B850 Considering Coulomb Couplings . . . . .	98
5.10	MM MD simulation and LC-DFTB: Absorption Spectra Shift of B800 und B850 Considering Coulomb Couplings . . . . .	98
6.1	Geometric Criteria of Hydrogen Bonds . . . . .	109
6.2	QM/MM, P <sub>1a</sub> : Retinal Orientation and H-Bonds Pattern . . . . .	111
6.3	QM/MM, P <sub>1a</sub> : Steric Clash in the Retinal Binding Pocket . . . . .	114
6.4	QM/MM, P <sub>1a</sub> : H-bond Pattern DC Gate . . . . .	116
6.5	QM/MM, P <sub>1a</sub> : E90 Orientation in the Homology Model . . . . .	117
6.6	QM/MM, P <sub>1a</sub> : H-bond Pattern of E90 found in X-ray . . . . .	118
6.7	QM/MM, P <sub>1a</sub> : H-bond Pattern E123 and D253 . . . . .	119
6.8	QM/MM, P <sub>1a</sub> : H-bond Pattern E123, D253 and K93 . . . . .	120
6.9	Pre-Gating Process . . . . .	122
6.10	Possible Deprotonation Pathway . . . . .	123
6.11	Turn of C <sub>13</sub> -Methyl Group of Retinal towards C128 . . . . .	124
6.12	Distances between D156–C128 and C128–Retinal . . . . .	124
6.13	MM MD, P <sub>1</sub> : H-bond Pattern RSBH <sup>+</sup> . . . . .	125
6.14	Homology Model: Water-Pore Formation . . . . .	129
6.15	X-ray Structure: Water-Pore Formation . . . . .	129
6.16	QM/MM, P <sub>1b</sub> : H-bond Pattern DC Gate . . . . .	131
6.17	QM/MM, P <sub>1a</sub> : Vibration Spectra of DC Gate Pattern . . . . .	136
6.18	QM/MM, P <sub>1b</sub> : Vibration Spectra of DC Gate Pattern . . . . .	137
A.1	BChl a, Measured Bond Lengths . . . . .	147
A.2	Simulated Absorption Spectra of ppR . . . . .	149
A.3	Binding Pocket of Retinal: HM Model versus X-ray . . . . .	151
A.4	Dihedral Angle Retinal . . . . .	153

A.5	RMSD-Helix A	154
A.6	RMSD-Helix B	154
A.7	RMSD-Helix C	154
A.8	RMSD-Helix D	155
A.9	RMSD-Helix E	155
A.10	RMSD-Helix F	155
A.11	RMSD-Helix G	156



---

## List of Tables

2.1	Possible Excitations of Organic Molecules . . . . .	8
4.1	Retinal: Environmental Effects on the Excitation Energies . . . . .	66
4.2	Supermolecule Calculation of Truncated BChl a Dimers: Different Distances .	73
4.3	Supermolecule Calculation of Truncated BChl a Dimers: Different Distances and Rotation . . . . .	73
4.4	Coulomb Couplings of Truncated BChl a Dimers . . . . .	74
4.5	Supermolecule Calculations and Coulomb Couplings: B850 Ring of the LH2 Complex . . . . .	75
5.1	Excitation Energies of QM/MM Optimized bR and ppR . . . . .	86
5.2	Absorption Maxima of bR and ppR Sampled by QM/MM MD Simulations .	87
5.3	Excitation Energies of the QM/MM Optimized LH2 Complex . . . . .	90
5.4	Coulomb Couplings of the QM/MM Optimized LH2 Complex . . . . .	91
5.5	LC-DFTB and ZINDO/S: Absorption Maxima, Environmental Effects, LH2 Complex . . . . .	96
5.6	LC-DFTB and ZINDO/S: Absorption Maxima, Environmental Effects and Coulomb Couplings, LH2 Complex . . . . .	97
6.1	Start Structures for the Study of P <sub>1a</sub> Based on X-ray . . . . .	107
6.2	Start Structures for the Study of P <sub>1b</sub> Based on X-ray . . . . .	107
6.3	QM Regions of P <sub>1a</sub> QM/MM MD Simulations . . . . .	108
6.4	QM/MM, P <sub>1a</sub> : H-bond Pattern of RSBH <sup>+</sup> and Dihedral Angle C <sub>12</sub> -C <sub>13</sub> =C <sub>14</sub> - C <sub>15</sub> of Retinal . . . . .	112
6.5	QM/MM, P <sub>1a</sub> : H-bond Percentage of the DC Gate . . . . .	115
6.6	QM/MM, P <sub>1a</sub> : H-bond Pattern of E90 to E123 and D253 . . . . .	118
6.7	QM/MM, P <sub>1a</sub> : E123 Orientation . . . . .	120
6.8	MM MD, P <sub>1</sub> : E123 Orientation . . . . .	127
6.9	QM/MM, P <sub>1b</sub> : H-bond Pattern E90 to E123 and D253 . . . . .	132
6.10	QM/MM, P <sub>1b</sub> : E123 Orientation . . . . .	133
6.11	Absorption Maxima of the P <sub>1</sub> State of ChR2-WT . . . . .	134

A.1	Vacuum Excitation Energies of all- <i>trans</i> Retinal . . . . .	143
A.2	Influence on Excitation Energies of the $\beta$ -Ionone Ring Twist Angle of 6-s-cis-11-cis PSB . . . . .	144
A.3	Part A: Previous Studies on Exciton Couplings of the LH2 Complex . . . . .	145
A.4	Part B: Previous Studies on Exciton Couplings of the LH2 Complex . . . . .	146
A.5	Excitation Energies of BChl a in Vacuum . . . . .	147
A.6	Bond Lengths of BChl a in Vacuum Optimized . . . . .	148
A.7	Dihedral Angle of the BChl a in Vacuum Optimized . . . . .	148
A.8	BLA of the BChl a Chromophores of the LH2 Complex and Excitation Energy Shift (Protein Environment–Vacuum) . . . . .	150
A.9	QM/MM Study of P1a: H-bond Percentage of the DC Gate . . . . .	152
A.10	QM/MM Study of P1a: H-bond Percentage of the Central Gate of the Homology Model . . . . .	153

---

## References

- [1] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, and E. Lindahl. GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX*, 1-2:19–25, 2015.
- [2] R. Ahlrichs, M. Bär, M. Häser, H. Horn, and C. Kölmel. Electronic structure calculations on workstation computers: The program system turbomole. *Chem. Phys. Lett.*, 162(3):165–169, 1989.
- [3] S. Ahuja, M. Eilers, A. Hirshfeld, E. C. Y. Yan, M. Ziliox, T. P. Sakmar, M. Sheves, and S. O. Smith. 6-s-cis Conformation and Polar Binding Pocket of the Retinal Chromophore in the Photoactivated State of Rhodopsin . *J. Am. Chem. Soc.*, 131(42):15160–15169, 2009.
- [4] A. Albini. *Photochemistry: Past, Present and Future*. Springer, Berlin, Heidelberg, 1 edition, 2016.
- [5] R. G. Alden, E. Johnson, V. Nagarajan, and W. W. Parson. Calculations of Spectroscopic Properties of the LH2 Bacteriochlorophyll-Protein Antenna Complex from Rhodospseudomonas Acidophila. *J. Phys. Chem. B*, 101(23):4667–4680, 1997.
- [6] J. Almlöf, T. H. Fischer, P. G. Gassman, A. Ghosh, and M. Häser. Electron Correlation in Tetrapyrroles. Ab initio Calculations on Porphyrin and the Tautomers of Chlorin. *J. Phys. Chem.*, 97(42):10964–10970, 1993.
- [7] T. Alshuth and M. Stockburger. Structural Changes in the Retinal Chromophore of Bacteriorhodopsin Studied by Resonance Raman Spectroscopy. *Ber. Bunsenges. Phys. Chem.*, 85:484–489, 1981.
- [8] A. Altun, S. Yokoyama, and K. Morokuma. Mechanism of Spectral Tuning Going from Retinal in Vacuo to Bovine Rhodopsin and its Mutants: Multireference ab Initio Quantum Mechanics/Molecular Mechanics Studies. *J. Phys. Chem. B*, 112(51):16883–16890, dec 2008.
- [9] A. Anda, L. De Vico, and T. Hansen. Intermolecular Modes between LH2 Bacteriochlorophylls and Protein Residues: The Effect on the Excitation Energies. *J. Phys. Chem. B*, 121(22):5499–5508, 2017.

- [10] A. Anda, T. Hansen, and L. De Vico. Multireference Excitation Energies for Bacteriochlorophylls A within Light Harvesting System 2. *J. Chem. Theory Comput.*, 12(3):1305–1313, 2016.
- [11] L. H. Andersen, I. B. Nielsen, M. B. Kristensen, M. O. El Ghazaly, S. Haacke, M. B. Nielsen, and M. Å. Petersen. Absorption of Schiff-Base Retinal Chromophores in Vacuo. *J. Am. Chem. Soc.*, 127(35):12347–12350, 2005.
- [12] A. J. A. Aquino, M. Barbatti, and H. Lischka. Excited-State Properties and Environmental Effects for Protonated Schiff Bases: A Theoretical Study. *ChemPhysChem*, 7(10):2089–2096, 2006.
- [13] A. Ardevol and G. Hummer. Retinal Isomerization and Water-Pore Formation in Channelrhodopsin-2. *Proc. Natl. Acad. Sci. U.S.A.*, 115(14):3557–3562, 2018.
- [14] G. Babitzki, G. Mathias, and P. Tavan. The Infrared Spectra of the Retinal Chromophore in Bacteriorhodopsin Calculated by a DFT/MM Approach. *J. Phys. Chem. B*, 113(30):10496–10508, 2009.
- [15] A. D. Bacon and M. C. Zerner. An Intermediate Neglect of Differential Overlap Theory for Transition Metal Complexes: Fe, Co and Cu Chlorides. *Theor. Chim. Acta*, 53(1):21–54, 1979.
- [16] R. Baer and D. Neuhauser. Density Functional Theory with Correct Long-Range Asymptotic Behavior. *Phys. Rev. Lett.*, 94(4):2–5, 2005.
- [17] C. Bamann, T. Kirsch, G. Nagel, and E. Bamberg. Spectral Characteristics of the Photocycle of Channelrhodopsin-2 and Its Implication for Channel Function. *J. Mol. Biol.*, 375(3):686–694, 2008.
- [18] R. J. Bartlett and M. Musiał. Coupled-Cluster Theory in Quantum Chemistry. *Rev. Mod. Phys.*, 79(1):291–352, 2007.
- [19] A. D. Becke. Density-Functional Exchange-Energy Approximation with Correct Asymptotic Behavior. *Phys. Rev. A*, 38(6), 1988.
- [20] J. Becker-Baldus, C. Bamann, K. Saxena, H. Gustmann, L. J. Brown, R. C. D. Brown, C. Reiter, E. Bamberg, J. Wachtveitl, H. Schwalbe, and C. Glaubitz. Enlightening the Photoactive Site of Channelrhodopsin-2 by DNP-Enhanced Solid-State NMR Spectroscopy. *Proc. Natl. Acad. Sci. U.S.A.*, 112(32):9896–9901, 2015.
- [21] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. Di Nola, and J. R. Haak. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.*, 81(8):3684–3690, 1984.

- 
- [22] P. H. Berens, S. R. White, and K. R. Wilson. Molecular Dynamics and Spectra. II. Diatomic Raman. *J. Chem. Phys.*, 75(2):515–529, 1981.
- [23] A. Berndt, O. Yizhar, L. A. Gunaydin, P. Hegemann, and K. Deisseroth. Bi-Stable Neural State Switches. *Nat. Neurosci.*, 12(2):229–234, 2009.
- [24] R. R. Birge and C. F. Zhang. Two-Photon Double Resonance Spectroscopy of Bacteriorhodopsin. Assignment of the Electronic and Dipolar Properties of the Low-Lying  $1\text{Ag}^*\text{-g}$ -like and  $1\text{B}^*\text{+ u}$ -like  $\pi$ ,  $\pi^*$  States. *J. Chem. Phys.*, 92(12):7178–7195, 1990.
- [25] A.-N. Bondar, M. Elstner, S. Suhai, J. C. Smith, and S. Fischer. Mechanism of Primary Proton Transfer in Bacteriorhodopsin. *Structure*, 12:1281–1288, 2004.
- [26] E. S. Boyden, F. Zhang, E. Bamberg, G. Nagel, and K. Deisseroth. Millisecond-Timescale, Genetically Targeted Optical Control of Neural Activity. *Nat. Neurosci.*, 8(9):1263–1268, 2005.
- [27] T. L. Brack and G. H. Atkinson. Picosecond Time-Resolved Resonance Raman Spectrum of the K-590 Intermediate in the Room Temperature Bacteriorhodopsin Photocycle. *J. Mol. Struct.*, 214:289–303, 1989.
- [28] M. S. Braiman, T. Mogi, T. Marti, L. J. Stern, H. G. Khorana, and K. J. Rothschild. Vibrational Spectroscopy of Bacteriorhodopsin Mutants: Light-Driven Proton Transport Involves Protonation Changes of Aspartic Acid Residues 85,96, and 212. *Phys. Rev. A*, 27:8516–8520, 1988.
- [29] B. R. Brooks, C. L. Brooks, A. D. Mackerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caffisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York, and M. Karplus. CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.*, 30:1545–1614, 2009.
- [30] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comput. Chem.*, 4(2):187–217, 1983.
- [31] E. Brunk and U. Rothlisberger. Mixed Quantum Mechanical/Molecular Mechanical Molecular Dynamics Simulations of Biological Systems in Ground and Electronically Excited States. *Chem. Rev.*, 115(12):6217–6263, 2015.
- [32] S. Bruun, D. Stoepler, A. Keidel, U. Kuhlmann, M. Luck, A. Diehl, M.-a. Geiger, D. Woodmansee, D. Trauner, P. Hegemann, H. Oschkinat, P. Hildebrandt, and K. Stehfest. Light - Dark Adaptation of Channelrhodopsin Involves Photoconversion between the all- trans and 13- cis Retinal Isomers. *Biochemistry*, 54:5389–5400, 2015.

- [33] Y. Cao, G. Váró, A. L. Klinger, D. M. Czajkowsky, M. S. Braiman, R. Needleman, and J. K. Lanyi. 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*, 32:1981–1990, 1993.
- [34] M. E. Casida. *Recent Advances in Density Functional Methods, Part I*. World Scientific, Singapur, 1995.
- [35] M. E. Casida and M. Huix-Rotllant. Progress in Time-Dependent Density-Functional Theory. *Annu. Rev. Phys. Chem.*, 63:287–323, 2012.
- [36] J. D. Chai and M. Head-Gordon. Systematic Optimization of Long-Range Corrected Hybrid Density Functionals. *J. Chem. Phys.*, 128(8):0–15, 2008.
- [37] C. Cheng, M. Kamiya, M. Takemoto, R. Ishitani, O. Nureki, N. Yoshida, and S. Hayashi. An Atomistic Model of a Precursor State of Light-Induced Channel Opening of Channelrhodopsin. *Biophys. J.*, 115(7):1281–1291, 2018.
- [38] I. Chizhov, G. Schmies, R. Seidel, J. R. Sydor, B. Lüttenberg, and M. Engelhard. The Photophobic Receptor from *Natronobacterium pharaonis*: Temperature and pH Dependencies of the Photocycle of Sensory Rhodopsin II. *Biophys. J.*, 75:999–1009, 1998.
- [39] M. G. Cory, M. C. Zerner, X. Hu, and K. Schulten. Electronic Excitations in Aggregates of Bacteriochlorophylls. *J. Phys. Chem. B*, 102(39):7640–7650, 1998.
- [40] C. J. Cramer. *Essentials to Computational Chemistry*. John Wiley & Sons, Chichester, 2 edition, 2004.
- [41] R. Croce, R. van Grondelle, H. van Amerongen, and I. van Stokkum. *Light Harvesting in Photosynthesis*. CRC press, 1 edition, 2018.
- [42] L. Cupellini, S. Jurinovich, M. Campetella, S. Caprasecca, C. a. Guido, S. M. Kelly, A. T. Gardiner, R. Cogdell, and B. Mennucci. An Ab Initio Description of the Excitonic Properties of LH2 and Their Temperature Dependence. *J. Phys. Chem. B*, 120(44):11348–11359, 2016.
- [43] C. Curutchet and B. Mennucci. Quantum Chemical Studies of Light Harvesting. *Chem. Rev.*, 117(2):294–343, 2017.
- [44] A. K. Dioumaev. Infrared Methods for Monitoring the Protonation State of Carboxylic Amino Acids in the Photocycle of Bacteriorhodopsin. *Biochemistry*, 66(11):1269–1276, 2001.

- 
- [45] P. Dohmen. Qm/mm study for the calculation of uv/vis absorption spectra of the light-harvesting complex iii. Master's thesis, Karlsruhe Institute of Technology (KIT), Karlsruhe, 2018.
- [46] A. Dreuw and M. Head-gordon. Single-Reference ab Initio Methods for the Calculation of Excited States of Large Molecules. *Chem. Rev.*, 105:4009–4037, 2005.
- [47] L. L. Duan, Y. Mei, Q. G. Zhang, and J. Z. Zhang. Intra-Protein Hydrogen Bonding is Dynamically Stabilized by Electronic Polarization. *J. Chem. Phys.*, 130(11), 2009.
- [48] A. Dwyer. *Weak Interactions and Excited States from Coulomb-Attenuated DFT*. PhD thesis, Durham University, 2011.
- [49] M. Eichinger, P. Tavan, J. Hutter, and M. Parrinello. A Hybrid Method for Solutes in Complex Solvents: Density Functional Theory Combined with Empirical Force Fields. *J. Chem. Phys.*, 110(21):10452–10467, 1999.
- [50] K. Eisenhauer, J. Kuhne, E. Ritter, A. Berndt, S. Wolf, E. Freier, F. Bartl, P. Hegemann, and K. Gerwert. In Channelrhodopsin-2 Glu-90 Is Crucial for Ion Selectivity and Is Deprotonated during the Photocycle. *J. Biol. Chem.*, 287(9):6904–6911, 2012.
- [51] M. Elstner, Q. Cui, P. Munih, E. Kaxiras, T. Frauenheim, and M. Karplus. Modeling Zinc in Biomolecules with the Self Consistent Charge-Density Functional Tight Binding (SCC-DFTB) Method: Applications to Structural and Energetic Analysis. *J. Comp. Chem.*, 24(5):565–581, 2003.
- [52] M. Elstner, T. Frauenheim, E. Kaxiras, G. Seifert, and S. Suhai. A Self-Consistent Charge Density-Functional Based Tight-Binding Scheme for Large Biomolecules. *Phys. Stat. Sol. B*, 217:357–376, 2000.
- [53] M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, T. Frauenheim, S. Suhai, and G. Seifert. Self-Consistent-Charge Density-Functional Tight-Binding Method for Simulations of Complex Materials Properties. *Phys. Rev. B*, 58(11):7260–7268, 1998.
- [54] O. P. Ernst, D. T. Lodowski, M. Elstner, P. Hegemann, and L. S. Brown. Microbial and Animal Rhodopsins : Structures , Functions , and Molecular Mechanisms. *Chem. Rev.*, 114(1):126–163, 2014.
- [55] R. E. Fenna and B. W. Matthews. Chlorophyll Arrangement in a Bacteriochlorophyll protein from Chlorohium limicola. *Nature*, 258:573–577, 1975.
- [56] N. Ferré and M. Olivucci. Probing the Rhodopsin Cavity with Reduced Retinal Models at the CASPT2/CASSCF/AMBER Level of Theory. *J. Am. Chem. Soc.*, 125(23):6868–6869, 2003.

- [57] W. B. Fischer, S. Sonar, T. Marti, H. Gobind Khorana, and K. J. Rothschild. Detection of a Water Molecule in the Active-Site of Bacteriorhodopsin: Hydrogen Bonding Changes during the Primary Photoreaction. *Biochemistry*, 33(43):12757–12762, 1994.
- [58] N. Foloppe and A. D. Mackerell. Acids: I. Parameter Optimization Based on Small Molecule and Condensed Phase Macromolecular Target Data. *J. Comput. Chem.*, 21(2):86–104, 2000.
- [59] J. S. Frähmcke, M. Wanko, P. Phatak, M. A. Mrogiński, and M. Elstner. The Protonation State of Glu181 in Rhodopsin Revisited: Interpretation of Experimental Data on the Basis of QM/MM Calculations. *J. Phys. Chem. B*, 114(34):11338–11352, 2010.
- [60] H. Frauenfelder, G. Chen, J. Berendzen, P. W. Fenimore, H. Jansson, B. H. McMahon, I. R. Stroe, J. Swenson, and R. D. Young. A Unified Model of Protein Dynamics. *Proc. Natl. Acad. Sci. U.S.A.*, 106(13):5129–5134, 2009.
- [61] F. Furche and R. Ahlrichs. Adiabatic Time-Dependent Density Functional Methods for Excited State Properties. *J. Chem. Phys.*, 117(16):7433–7447, 2002.
- [62] J. Gao, W.-J. Shi, J. Ye, X. Wang, H. Hirao, and Y. Zhao. QM/MM Modeling of Environmental Effects on Electronic Transitions of the FMO Complex. *J. Phys. Chem. B*, 117(13):3488–3495, 2013.
- [63] M. Gaus, Q. Cui, and M. Elstner. DFTB3: Extension of the Self-Consistent-Charge Density-Functional Tight-Binding Method (SCC-DFTB). *J. Chem. Theory Comput.*, 7(4):931–948, 2011.
- [64] M. Gaus, A. Goez, and M. Elstner. Parametrization and Benchmark of DFTB3 for Organic Molecules. *J. Chem. Theory Comput.*, 9(1):338–354, 2013.
- [65] M. Gaus, X. Lu, M. Elstner, and Q. Cui. Parameterization of DFTB3/3OB for Sulfur and Phosphorus for Chemical and Biological Applications. *J. Chem. Theory Comput.*, 10(4):1518–1537, 2014.
- [66] S. Georgakopoulou, R. N. Frese, E. Johnson, C. Koolhaas, R. J. Cogdell, R. Van Grondelle, and G. Van Der Zwan. Absorption and CD Spectroscopy and Modeling of Various LH2 Complexes from Purple Bacteria. *Biophys. J.*, 82(4):2184–2197, 2002.
- [67] K. Gerwert, E. Freier, and S. Wolf. The Role of Protein-Bound Water Molecules in Microbial Rhodopsins. *Biochim. Biophys. Acta*, 1837(5):606–613, 2014.
- [68] K. Gerwert, G. Souvignier, and B. Hess. 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. Natl. Acad. Sci. USA*, 87(24):9774–9778, 1990.

- 
- [69] M. Gouterman. Spectra of Porphyrins. *J. Mol. Spectrosc.*, 6(C):138–163, 1961.
- [70] M. Gouterman and G. H. Wagnière. Spectra of Porphyrins. *J. Mol. Spectrosc.*, 127:108–127, 1963.
- [71] D. Gradmann, A. Berndt, F. Schneider, and P. Hegemann. Rectification of the Channelrhodopsin Early Conductance. *Biophys. J.*, 101(5):1057–1068, 2011.
- [72] L. A. Gunaydin, O. Yizhar, A. Berndt, V. S. Sohal, K. Deisseroth, and P. Hegemann. Ultrafast Optogenetic Control. *Nat. Neurosci.*, 13(3):387–392, 2010.
- [73] Y. Guo, F. E. Beyle, B. M. Bold, H. C. Watanabe, A. Koslowski, W. Thiel, P. Hegemann, M. Marazzi, and M. Elstner. Active Site Structure and Absorption Spectrum of Channelrhodopsin-2 Wild-Type and C128T Mutant. *Chem. Sci.*, 7(6):3879–3891, 2016.
- [74] Y. Guo, F. E. Wolff, I. Schapiro, M. Elstner, and M. Marazzi. Different Hydrogen Bonding Environments of the Retinal Protonated Schiff Base Control the Photoisomerization in Channelrhodopsin-2. *Phys. Chem. Chem. Phys.*, 20(43):27501–27509, 2018.
- [75] H. Haken and H. C. Wolf. *Molekülphysik und Quantenchemie: Einführung in die Experimentellen und Theoretischen Grundlagen*. Springer-Lehrbuch. Springer, Berlin, 4 edition, 2003.
- [76] C. Hättig. *Structure Optimizations for Excited States with Correlated Second-Order Methods: CC2 and ADC(2)*, volume 50 of *Advances in Quantum Chemistry*. Academic Press, 2005.
- [77] C. Hättig and F. Weigend. Cc2 excitation energy calculations on large molecules using the resolution of the identity approximation. *J. Chem. Phys.*, 113(13):5154–5161, 2000.
- [78] T. Helgaker, J. Poul, and J. Olsen. *Molecular Electronic-Structure Theory*. Wiley, Chichester, 1 edition, 2012.
- [79] R. Henderson and P. N. T. Unwin. Three-Dimensional Model of Purple Membrane obtained by Electron Microscopy. *Nature*, 257:28–32, 1975.
- [80] B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.*, 4(3):435–447, 2008.
- [81] M. Higashi, T. Kosugi, S. Hayashi, and S. Saito. Theoretical Study on Excited States of Bacteriochlorophyll a in Solutions with Density Functional Assessment. *J. Phys. Chem. B*, 118(37):10906–10918, 2014.

- [82] M. Hoffmann, M. Wanko, P. Strodel, P. H. König, T. Frauenheim, K. Schulten, W. Thiel, E. Tajkhorshid, and M. Elstner. Color Tuning in Rhodopsins: the Mechanism for the Spectral Shift between Bacteriorhodopsin and Sensory Rhodopsin II. *J. Am. Chem. Soc.*, 128(33):10808–10818, 2006.
- [83] W. G. Hoover. Canonical Dynamics: Equilibrium Phase-Space Distributions. *Phys. Rev. A*, 31(3):1695–1697, 1985.
- [84] J. Huang and A. D. MacKerell. CHARMM36 All-Atom Additive Protein Force Field: Validation Based on Comparison to NMR Data. *J. Comput. Chem.*, 34(25):2135–2145, 2013.
- [85] W. Humphrey, A. Dalke, and K. Schulten. VMD – Visual Molecular Dynamics. *J. Mol. Graph.*, 14:33–38, 1996.
- [86] S. Höfener. Personal communication, 2018.
- [87] A. Inaguma, H. Tsukamoto, H. E. Kato, T. Kimura, T. Ishizuka, S. Oishi, H. Yawo, O. Nureki, and Y. Furutani. Chimeras of Channelrhodopsin-1 and-2 from *Chlamydomonas Reinhardtii* Exhibit Distinctive Light-Induced Structural Changes from Channelrhodopsin-2. *J. Biol. Chem.*, 290(18):11623–11634, 2015.
- [88] S. Ito, H. E. Kato, R. Taniguchi, T. Iwata, O. Nureki, and H. Kandori. Water-Containing Hydrogen-Bonding Network in the Active Center of Channelrhodopsin. *J. Am. Chem. Soc.*, 136(9):3475–3482, 2014.
- [89] D. Jacquemin, E. A. Perpète, G. Scalmani, M. J. Frisch, R. Kobayashi, and C. Adamo. Assessment of the Efficiency of Long-Range Corrected Functionals for Some Properties of Large Compounds. *J. Chem. Phys.*, 126(14):0–12, 2007.
- [90] D. Jacquemin, E. A. Perpète, G. E. Scuseria, I. Ciofini, and C. Adamo. TD-DFT Performance for the visible Absorption Spectra of Organic Dyes: Conventional versus Long-Range Hybrids. *J. Chem. Theory Comput.*, 4(1):123–135, 2008.
- [91] S. J. Jang and B. Mennucci. Delocalized Excitons in Natural Light-Harvesting Complexes. *Rev. Mod. Phys.*, 90(3):35003, 2018.
- [92] C. L. Janssen and I. M. Nielsen. New diagnostics for coupled-cluster and moller–plesset perturbation theory. *Chem. Phys. Lett.*, 290(4):423 – 430, 1998.
- [93] F. Jensen. *Introduction to Computational Chemistry*. John Wiley & Sons, Chichester, 2 edition, 2007.
- [94] X. Jia, Y. Mei, J. Z. H. Zhang, and Y. Mo. Hybrid QM/MM study of FMO Complex with Polarized Protein-Specific Charge. *Sci. Rep.*, 5:1–10, 2015.

- 
- [95] W. L. Jorgensen, J. Chandrasekhar, and J. D. Madura. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.*, 79(2):926–935, 1983.
- [96] S. Jurinovich, L. Cupellini, C. A. Guido, and B. Mennucci. Exat: Excitonic Analysis Tool. *J. Comput. Chem.*, 39(5):279–286, 2018.
- [97] S. Jurinovich, L. Viani, C. Curutchet, and B. Mennucci. Limits and Potentials of Quantum Chemical Methods in Modelling Photosynthetic Antennae. *Phys. Chem. Chem. Phys.*, 17(46):30783–30792, 2015.
- [98] S. Kaminski, M. Gaus, P. Phatak, D. von Stetten, M. Elstner, and M. A. Mrogiński. Vibrational Raman Spectra from the Self-Consistent Charge Density Functional Tight Binding Method via Classical Time-Correlation Functions. *J. Chem. Theory Comput.*, 6(4):1240–1255, 2010.
- [99] H. Kandori, K. Shimono, Y. Sudo, M. Iwamoto, Y. Shichida, and N. Kamo. Structural Changes of pharaonis Phoborhodopsin upon Photoisomerization of the Retinal Chromophore: Infrared Spectral Comparison with Bacteriorhodopsin. *Biochemistry*, 40(31):9238–9246, 2001.
- [100] H. E. Kato, F. Zhang, O. Yizhar, C. Ramakrishnan, T. Nishizawa, K. Hirata, J. Ito, Y. Aita, T. Tsukazaki, S. Hayashi, P. Hegemann, A. D. Maturana, R. Ishitani, K. Deiseroth, and O. Nureki. Crystal Structure of the Channelrhodopsin Light-Gated Cation Channel. *Nature*, 482(7385):369–374, 2012.
- [101] E. P. Kenny and I. Kassal. Benchmarking Calculations of Excitonic Couplings between Bacteriochlorophylls. *J. Phys. Chem. B*, 120(1):25–32, 2016.
- [102] H. G. Khorana, G. E. Gerber, W. C. Herlihy, C. P. Gray, R. J. Andereg, K. Nihei, and K. Biemann. Amino Acid Sequence of Bacteriorhodopsin. *Proc. Nat. Acad. Sci. USA*, 76(10):5046–5050, 1979.
- [103] P. Klàn and J. Wirz. *Photochemistry of Organic Compounds: From Concepts to Practice*. Postgraduate chemistry series. Wiley, Chichester, 2009.
- [104] J. B. Klauda, R. M. Venable, J. A. Freites, J. W. O’Connor, D. J. Tobias, C. Mondragon-Ramirez, I. Vorobyov, A. D. MacKerell, and R. W. Pastor. Update of the CHARMM All-Atom Additive Force Field for Lipids: Validation on Six Lipid Types. *J. Phys. Chem. B*, 114(23):7830–7843, 2010.
- [105] U. Kleinekathöfer. Personal communication, 2017.
- [106] R. S. Knox and B. Q. Spring. Dipole Strengths in the Chlorophylls. *Photochem. Photobiol.*, 77(5):497–501, 2003.

- [107] W. Koch and M. C. Holthausen. *A Chemists Guide to Density Functional Theory*. John Wiley & Sons, Weinheim, 2 edition, 2008.
- [108] G. G. Kochendoerfer, S. W. Lin, T. P. Sakmar, and R. A. Mathies. How Color Visual Pigments are Tuned. *Trends Biochem. Sci.*, 24(8):300–305, 1999.
- [109] J. Koepke, X. Hu, C. Muenke, K. Schulten, and H. Michel. The Crystal Structure of the Light-Harvesting Complex II (B800-850) from Rhodospirillum Molischianum. *Structure*, 4(5):581–597, 1996.
- [110] W. Kohn and L. J. Sham. Self-Consistent Equation Including Exchange and Correlation Effects. *Phys. Rev. A*, 140(4), 1965.
- [111] C. König and J. Neugebauer. Quantum Chemical Description of Absorption Properties and Excited-State Processes in Photosynthetic Systems. *ChemPhysChem*, 13(2):386–425, 2012.
- [112] T. Körzdörfer, R. M. Parrish, J. S. Sears, C. D. Sherrill, and J.-L. Brédas. On the relationship between bond-length alternation and many-electron self-interaction error. *J. Chem. Phys.*, 137(12):124305, 2012.
- [113] A. Kosłowski, M. E. Beck, and W. Thiel. Implementation of a General Multireference Configuration Interaction Procedure with Analytic Gradients in a Semiempirical Context using the Graphical Unitary Group Approach. *J. Comput. Chem.*, 6:714–726, 2003.
- [114] J. J. Kranz and M. Elstner. Simulation of Singlet Exciton Diffusion in Bulk Organic Materials. *J. Chem. Theo. Comp.*, 12(9):4209–4221, 2016.
- [115] J. J. Kranz, M. Elstner, B. Aradi, T. Frauenheim, V. Lutsker, A. D. Garcia, and T. A. Niehaus. Time-Dependent Extension of the Long-Range Corrected Density Functional Based Tight-Binding Method. *J. Chem. Theo. Comp.*, 13(4):1737–1747, 2017.
- [116] N. Krause, C. Engelhard, J. Heberle, R. Schlesinger, and R. Bittl. Structural Differences between the Closed and Open States of Channelrhodopsin-2 as Observed by EPR Spectroscopy. *FEBS Lett.*, 587(20):3309–3313, 2013.
- [117] B. P. Krueger, G. D. Scholes, and G. R. Fleming. Calculation of Couplings and Energy-Transfer Pathways between the Pigments of LH2 by the ab Initio Transition Density Cube Method. *J. Phys. Chem. B*, 102(27):5378–5386, 1998.
- [118] T. Kubař, K. Welke, and G. Groenhof. New QM/MM Implementation of the DFTB3 Method in the Gromacs Package. *J. Comput. Chem.*, 36(26):1978–1989, 2015.
- [119] T. Kubař, P. Benjamin Woiczikowski, G. Cuniberti, and M. Elstner. Efficient Calculation of Charge-Transfer Matrix Elements for Hole Transfer in DNA. *J. of Phys. Chem. B*, 112(26):7937–7947, 2008.

- 
- [120] J. Kuhne, K. Eisenhauer, E. Ritter, P. Hegemann, K. Gerwert, and F. Bartl. Early Formation of the Ion-Conducting Pore in Channelrhodopsin-2. *Angew. Chem. Int. Ed.*, 54(16):4953–4957, 2015.
- [121] J. K. Lanyi. Bacteriorhodopsin. *Annu. Rev. Physiol.*, 66:665–688, 2004.
- [122] C. Lee, W. Yang, and R. G. Parr. Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Phys. Rev. A*, 37(2):785–789, 1988.
- [123] I. Leontyev and A. Stuchebrukhov. Accounting for Electronic Polarization in Non-Polarizable Force Fields. *Phys. Chem. Chem. Phys.*, 13:2613–2626, 2011.
- [124] X. Li, D. V. Gutierrez, M. G. Hanson, J. Han, M. D. Mark, H. Chiel, P. Hegemann, L. T. Landmesser, and S. Herlitze. Fast Noninvasive Activation and Inhibition of Neural and Network Activity by Vertebrate Rhodopsin and Green Algae Channelrhodopsin. *Proc. Natl. Acad. Sci. U.S.A.*, 102(49):17816–17821, 2005.
- [125] J. Linnanto and J. Korppi-Tommola. Quantum Chemical Simulation of Excited States of Chlorophylls, Bacteriochlorophylls and their Complexes. *Phys. Chem. Chem. Phys.*, 8(6):663–687, 2006.
- [126] N. H. List. Personal communication, 2018.
- [127] N. H. List, C. Curutchet, S. Knecht, B. Mennucci, and J. Kongsted. Toward Reliable Prediction of the Energy Ladder in Multichromophoric Systems: A Benchmark Study on the FMO Light-Harvesting Complex. *J. Chem. Theory Comput.*, 9(11):4928–4938, 2013.
- [128] R. S. H. Liu, E. Krogh, X.-Y. Li, D. Mead, L. U. Colmenares, J. R. Thiel, J. Ellis, D. Wong, and A. E. Asato. Analyzing the Red-Shift Characteristics of Azulenic, Naphthyl, Other Ring-Fused and Retinyl Pigment Analogs of Bacteriorhodopsin. *Photochem. Photobiol.*, 58(5):701–705, 1993.
- [129] E. Livshits and R. Baer. A Well-Tempered Density Functional Theory of Electrons in Molecules. *Phys. Chem. Chem. Phys.*, 9(23):2932–2941, 2007.
- [130] S. L. Logunov, L. Song, and M. A. El-Sayed. Excited-State Dynamics of a Protonated Retinal Schiff Base in Solution. *J. Phys. Chem.*, 100(47):18586–18591, 1996.
- [131] C. S. López, O. N. Faza, S. L. Estévez, and A. R. De Lera. Computation of Vertical Excitation Energies of Retinal and Analogs: Scope and Limitations. *J. Comput. Chem.*, 27(1):116–123, 2006.
- [132] V. A. Lórenz-Fonfría, C. Bamann, T. Resler, R. Schlesinger, E. Bamberg, and J. Heberle. Temporal Evolution of Helix Hydration in a Light-Gated Ion Channel Correlates with Ion Conductance. *Proc. Natl. Acad. Sci. U.S.A.*, page 201511462, 2015.

- [133] V. a. Lórenz-Fonfría and J. Heberle. Channelrhodopsin Unchained: Structure and Mechanism of a Light-Gated Cation Channel. *Biochim. Biophys. Acta*, 1837(5):626–642, 2014.
- [134] V. A. Lórenz-Fonfría, T. Resler, N. Krause, M. Nack, M. Gossing, G. Fischer von Mollard, C. Bamann, E. Bamberg, R. Schlesinger, and J. Heberle. Transient Protonation Changes in Channelrhodopsin-2 and their Relevance to Channel Gating. *Proc. Natl. Acad. Sci. U.S.A.*, 110(14):E1273–81, 2013.
- [135] V. a. Lórenz-Fonfría, B.-J. Schultz, T. Resler, R. Schlesinger, C. Bamann, E. Bamberg, and J. Heberle. Pre-Gating Conformational Changes in the ChETA Variant of Channelrhodopsin-2 Monitored by Nanosecond IR Spectroscopy. *J. Am. Chem. Soc.*, 137(5):1850–61, 2015.
- [136] K. Ludmann, C. Gergely, and G. Váró. Kinetic and Thermodynamic Study of the Bacteriorhodopsin Photocycle over a Wide pH Range. *Biophys. J.*, 214:3110–3119, 1989.
- [137] V. Lutsker, B. Aradi, and T. A. Niehaus. Implementation and Benchmark of a Long-Range Corrected Functional in the Density Functional Based Tight-Binding Method. *J. Chem. Phys.*, 143:184107, 2015.
- [138] A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, and M. Karplus. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B*, 102(18):3586–3616, 1998.
- [139] M. E. Madjet, A. Abdurahman, and T. Renger. Intermolecular Coulomb Couplings from ab initio Electrostatic Potentials: Application to Optical Transitions of Strongly Coupled Pigments in Photosynthetic Antennae and Reaction Centers. *J. Phys. Chem. B*, 110(34):17268–17281, 2006.
- [140] A. Maeda. Application of FTIR Spectroscopy to the Structural Study on the Function of Bacteriorhodopsin. *Isr. J. Chem.*, 35:387–400, 1995.
- [141] S. Maity. Personal communication, 2019.
- [142] R. Mathies and L. Stryer. Retinal has a Highly Dipolar Vertically Excited Singlet State: Implications of Vision. *Proc. Natl. Acad. Sci. USA*, 73(7):2169–2173, 1976.
- [143] R. A. Mathies, S. Lin, J. B. Ames, and W. T. Pollard. From Femtoseconds To Biology: Mechanism Of The Light-Driven Proton Pump In Bacteriorhodopsin. *Annu. Rev. Biophys. Biophys. Chem.*, 20:491–518, 1991.

- 
- [144] B. W. Matthews, R. E. Fenna, M. C. Bolognesi, and M. F. Schmid. Structure of a Bacteriochlorophyll *a*-Protein from the Green Photosynthetic Bacterium *Prosthecochloris aestuarii*. *J. Mol. Biol.*, 131:259–285, 1979.
- [145] V. May and O. Kühn. *Charge and Energy Transfer Dynamics in Molecular Systems*. Wiley-VCH, Weinheim, 2 edition, 2004.
- [146] G. McDermott, S. M. Prince, A. A. Freer, A. M. Hawthornthwaite-Lawless, M. Z. Papiz, R. J. Cogdell, and N. W. Isaacs. Crystal Structure of an Integral Membrane Light-Harvesting Complex from Photosynthetic Bacteria. *Nature*, 374:517–521, 1995.
- [147] M. T. Milder, B. Brüggemann, R. van Grondelle, and J. L. Herek. Revisiting the Optical Properties of the FMO Protein. *Photosyn. Res.*, 104(2):257–274, 2010.
- [148] A. Miller and D. Oesterhalt. Kinetic Optimization of Bacteriorhodopsin by Aspartic Acid 96 as an Internal Proton Donor. *Biochim. Biophys. Acta*, 1020:57–64, 1990.
- [149] A. Muñoz Losa, C. Curutchet, B. P. Krueger, L. R. Hartsell, and B. Mennucci. Fretting about FRET: Failure of the Ideal Dipole Approximation. *Biophys. J.*, 96(12):4779–4788, 2009.
- [150] V. Muders, S. Kerruth, V. A. Lórenz-Fonfría, C. Bamann, J. Heberle, and R. Schlesinger. Resonance Raman and FTIR Spectroscopic Characterization of the Closed and Open States of Channelrhodopsin-1. *FEBS lett.*, 588(14):2301–6, 2014.
- [151] M. Müller, C. Bamann, E. Bamberg, and W. Kühnbrandt. Projection Structure of Channelrhodopsin-2 at 6 Å Resolution by Electron Crystallography. *J. Mol. Biol.*, 414(1):86–95, 2011.
- [152] M. Nack, I. Radu, C. Bamann, E. Bamberg, and J. Heberle. The Retinal Structure of Channelrhodopsin-2 Assessed by Resonance Raman Spectroscopy. *FEBS Letters*, 583(22):3676–3680, 2009.
- [153] M. Nack, I. Radu, M. Gossing, C. Bamann, E. Bamberg, G. F. von Mollard, and J. Heberle. The DC gate in Channelrhodopsin-2: Crucial Hydrogen Bonding Interaction between C128 and D156. *Photochem. Photobiol. Sci.*, 9(2):194–198, 2010.
- [154] G. Nagel, M. Brauner, J. F. Liewald, N. Adeishvili, E. Bamberg, and A. Gottschalk. Light Activation of Channelrhodopsin-2 in Excitable Cells of *Caenorhabditis Elegans* Triggers Rapid Behavioral Responses. *Curr. Biol.*, 15(24):2279–2284, 2005.
- [155] G. Nagel, D. Ollig, M. Fuhrmann, S. Kateriya, A. M. Musti, E. Bamberg, and P. Hegemann. Channelrhodopsin-1: a Light-Gated Proton Channel in Green Algae. *Science*, 296(5577):2395–2398, 2002.

- [156] G. Nagel, T. Szellas, W. Huhn, S. Kateriya, N. Adeishvili, P. Berthold, D. Ollig, P. Hegemann, and E. Bamberg. Channelrhodopsin-2, a Directly Light-Gated Cation-Selective Membrane Channel. *Proc. Natl. Acad. Sci. U.S.A.*, 100(24):13940–5, 2003.
- [157] F. Neese. A Spectroscopy Oriented Configuration Interaction Procedure. *J. Chem. Phys.*, 119(18), 2003.
- [158] F. Neese. Software Update: the ORCA Program System, Version 4.0. *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 8(1):4–9, 2018.
- [159] J. Neugebauer. Photophysical Properties of Natural Light-Harvesting Complexes Studied by Subsystem Density Functional Theory. *J. Phys. Chem. B*, 112(7):2207–2217, 2008.
- [160] J. Neugebauer. Subsystem-Based Theoretical Spectroscopy of Biomolecules and Biomolecular Assemblies. *ChemPhysChem*, 10(18):3148–3173, 2009.
- [161] M. K. Neumann-Verhoeven, K. Neumann, C. Bamann, I. Radu, J. Heberle, E. Bamberg, and J. Wachtveitl. Ultrafast Infrared Spectroscopy on Channelrhodopsin-2 Reveals Efficient Energy Transfer from the Retinal Chromophore to the Protein. *J. Am. Chem. Soc.*, 135(18):6968–6976, 2013.
- [162] R. Neutze, E. Pebay-Peyroula, K. Edman, A. Royant, J. Navarro, and E. M. Landau. Bacteriorhodopsin: A High-Resolution Structural View of Vectorial Proton Transport. *Biochim. Biophys. Acta*, 1565(2):144–167, 2002.
- [163] B. Nie, J. Stutzman, and A. Xie. A Vibrational Spectral Maker for Probing the Hydrogen-Bonding Status of Protonated Asp and Glu Residues. *Biophys. J.*, 88(4):2833–47, 2005.
- [164] T. Niehaus. Approximate Time-Dependent Density Functional Theory. *J. Mol. Struct.: THEOCHEM*, 914(1-3):38–49, 2009.
- [165] T. A. Niehaus and F. Della Sala. Range Separated Functionals in the Density Functional Based Tight-Binding Method: Formalism. *Phys. Status Solidi B*, 249(2):237–244, 2012.
- [166] T. A. Niehaus, S. Suhai, F. Della Sala, M. Elstner, G. Seifert, and T. Frauenheim. Tight-Binding Approach to Time-Dependent Density-Functional Response Theory. *Phys. Rev. B Condens. Matter Mater. Phys.*, 63(8):1–9, 2001.
- [167] I. B. Nielsen, L. Lammich, and L. H. Andersen. S1 and S2 Excited States of Gas-Phase Schiff-Base Retinal Chromophores. *Phys. Rev. Lett.*, 96(1):13–16, 2006.
- [168] S. Nosé. A Molecular Dynamics Method for Simulations in the Canonical Ensemble. *Mol. Phys.*, 52(2):255–268, 1984.

- 
- [169] M. Nottoli, S. Jurinovich, L. Cupellini, A. T. Gardiner, R. Cogdell, and B. Mennucci. The Role of Charge-Transfer States in the Spectral Tuning of Antenna Complexes of Purple Bacteria. *Photosyn. Res.*, 137(2):215–226, 2018.
- [170] D. Oesterhelt and W. Stoeckenius. Functions of a New Photoreceptor Membrane. *Proc. Nat. Acad. Sci. USA*, 70(10):2853–2857, 1973.
- [171] C. Olbrich, T. L. C. Jansen, J. Liebers, M. Aghtar, J. Strümpfer, K. Schulten, J. Knoester, and U. Kleinekathöfer. From Atomistic Modeling to Excitation Transfer and Two-Dimensional Spectra of the FMO Light-Harvesting Complex. *J. Phys. Chem. B*, 115(26):8609–8621, 2011.
- [172] C. Olbrich and U. Kleinekathöfer. Time-Dependent Atomistic View on the Electronic Relaxation in Light-Harvesting System II. *J. Phys. Chem. B*, 114(38):12427–12437, 2010.
- [173] M. Parrinello and A. Rahman. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.*, 52(12):7182–7190, 1981.
- [174] W. W. Parson. *Modern Optical Spectroscopy: With Exercises and Examples from Biophysics and Biochemistry*. Springer, Berlin, Heidelberg, 2 edition, 2015.
- [175] J. P. Perdew, K. Burke, and M. Ernzerhof. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.*, 77(18):3865–3868, 1996.
- [176] M. Persico and G. Granucci. *Photochemistry: A Modern Theoretical Perspective*. Springer, Cham, 1 edition, 2018.
- [177] P. Phatak, J. S. Frähmcke, M. Wanko, M. Hoffmann, P. Strodel, J. C. Smith, S. Suhai, A.-N. Bondar, and M. Elstner. Long-Distance Proton Transfer with a Break in the Bacteriorhodopsin Active Site. *J. Am. Chem. Soc.*, 131(20):7064–7078, 2009.
- [178] P. Phatak, N. Ghosh, H. Yu, Q. Cui, and M. Elstner. Amino Acids with an Intermolecular Proton Bond as Proton Storage Site in Bacteriorhodopsin. *Proc. Natl. Acad. Sci. U.S.A.*, 105(50):19672–19677, 2008.
- [179] P. A. Plötz, T. Niehaus, and O. Kühn. A New Efficient Method for Calculation of Frenkel Exciton Parameters in Molecular Aggregates. *J. Chem. Phys.*, 140(17), 2014.
- [180] I. Radu, C. Bamann, M. Nack, G. Nagel, E. Bamberg, and J. Heberle. Conformational Changes of Channelrhodopsin-2. *J. Am. Chem. Soc.*, 131:7313–7319, 2009.
- [181] T. Resler, B.-J. Schultz, V. A. Lórenz-Fonfría, R. Schlesinger, and J. Heberle. Kinetic and Vibrational Isotope Effects of Proton Transfer Reactions in Channelrhodopsin-2. *Biophys. J.*, 109(2):287–297, 2015.

- [182] E. Ritter, P. Piwowarski, P. Hegemann, and F. J. Bartl. Light-Dark Adaptation of Channelrhodopsin C128T Mutant. *J. Biol. Chem.*, 288(15):10451–10458, 2013.
- [183] E. Ritter, K. Stehfest, A. Berndt, P. Hegemann, and F. J. Bartl. Monitoring Light-Induced Structural Changes of Channelrhodopsin-2 by UV-Visible and Fourier Transform Infrared Spectroscopy. *J. Biol. Chem.*, 283(50):35033–35041, 2008.
- [184] A. M. Rosnik and C. Curutchet. Theoretical Characterization of the Spectral Density of the Water-Soluble Chlorophyll-Binding Protein from Combined Quantum Mechanics/Molecular Mechanics Molecular Dynamics Simulations. *J. Chem. Theory Comput.*, 11(12):5826–5837, 2015.
- [185] I. V. Rostov, R. D. Amos, R. Kobayashi, G. Scalmani, and M. J. Frisch. Studies of the Ground and Excited-State Surfaces of the Retinal Chromophore using CAM-B3LYP. *J. Phys. Chem. B*, 114(16):5547–5555, 2010.
- [186] E. Runge and E. K. U. Gross. Density-Functional for Time-Dependent Systems. *Phys. Rev. Lett.*, 52(12):997–1000, 1984.
- [187] I. Schapiro and F. Neese. SORCI for Photochemical and Thermal Reaction Paths: A Benchmark Study. *Comput. Theor. Chem.*, 1040-1041:84–98, 2014.
- [188] M. Schmidt Am Busch, F. Müh, M. El-Amine Madjet, and T. Renger. The Eighth Bacteriochlorophyll Completes the Excitation Energy Funnel in the FMO Protein. *J. Phys. Chem. Lett.*, 2(2):93–98, 2011.
- [189] F. Schneider, C. Grimm, and P. Hegemann. Biophysics of Channelrhodopsin. *Annu. Rev. Biophys.*, 44:167–186, 2015.
- [190] G. D. Scholes, I. R. Gould, R. J. Cogdell, and G. R. Fleming. Ab Initio Molecular Orbital Calculations of Electronic Couplings in the LH2 Bacterial Light-Harvesting Complex of Rps. Acidophila. *J. Phys. Chem. B*, 103(13):2543–2553, 1999.
- [191] M. Schreiber and V. Buss. Origin of the Bathochromic Shift in the Early Photointermediates of the Rhodopsin Visual Cycle: A CASSCF/CASPT2 Study. *Int. J. Quantum Chem.*, 95(6):882–889, 2003.
- [192] F. Segatta, L. Cupellini, S. Jurinovich, S. Mukamel, M. Dapor, S. Taioli, M. Garavelli, and B. Mennucci. A Quantum Chemical Interpretation of Two-Dimensional Electronic Spectroscopy of Light-Harvesting Complexes. *J. Am. Chem. Soc.*, 139(22):7558–7567, 2017.
- [193] R. Send and D. Sundholm. Coupled-Cluster Studies of the Lowest Excited States of the 11-cis-Retinal Chromophore. *Phys. Chem. Chem. Phys.*, 9(22):2862–2867, 2007.

- 
- [194] R. Send and D. Sundholm. Stairway to the Conical Intersection: A Computational Study of the Retinal Isomerization. *J. Phys. Chem. A*, 111(36):8766–8773, 2007.
- [195] R. Send and D. Sundholm. The Role of the  $\beta$ -Ionone Ring in the Photochemical Reaction of Rhodopsin. *J. Phys. Chem. A*, 111(1):27–33, 2007.
- [196] R. Send, D. Sundholm, M. P. Johansson, and F. Pawłowski. Excited State Potential Energy Surfaces of Polyenes and Protonated Schiff Bases. *J. Chem. Theory Comput.*, 5:2401–2414, 2009.
- [197] H. M. Senn and W. Thiel. QM/MM Methods for Biomolecular Systems. *Angew. Chem. Int. Ed.*, 48:1198–1229, 2009.
- [198] M. Shibata, T. Tanimoto, and H. Kandori. Water Molecules in the Schiff Base Region of Bacteriorhodopsin. *J. Am. Chem. Soc.*, 125(44):13312–13313, 2003.
- [199] K. Shimono, Y. Furutani, H. Kandori, and N. Kamo. A Pharaonis Phoborhodopsin Mutant with the Same Retinal Binding Site Residues As in Bacteriorhodopsin. *Biochemistry*, 41(20):6504–6509, 2002.
- [200] M. R. Silva-Junior and W. Thiel. Benchmark of Electronically Excited States for Semiempirical Methods: MNDO, AM1, PM3, OM1, OM2, OM3, INDO/S, and INDO/S2. *J. Chem. Theory Comput.*, 6:1546–1564, 2010.
- [201] S. O. Smith, J. A. Pardoën, P. P. J. Mulder, B. Curry, J. Lugtenburg, and R. Mathies. Chromophore Structure in Bacteriorhodopsin’s O<sub>640</sub> Photointermediate. *Biochemistry*, 22:6141–6148, 1983.
- [202] J. L. Spudich, C. S. Yang, K. H. Jung, and E. N. Spudich. Retinylidene Proteins: Structures and Functions from Archaea to Humans. *Annu. Rev. Cell Dev. Biol.*, 16:365–392, 2000.
- [203] K. Stehfest, E. Ritter, A. Berndt, F. Bartl, and P. Hegemann. The Branched Photocycle of the Slow-Cycling Channelrhodopsin-2 Mutant C128T. *J. Mol. Biol.*, 398(5):690–702, 2010.
- [204] C. Steinmann and J. Kongsted. Electronic Energy Transfer in Polarizable Heterogeneous Environments: A Systematic Investigation of Different Quantum Chemical Approaches. *J. Chem. Theory Comput.*, 11(9):4283–4293, 2015.
- [205] R. E. Stratmann, G. E. Scuseria, and M. J. Frisch. An Efficient Implementation of Time-Dependent Density-Functional Theory for the Calculation of Excitation Energies of Large Molecules. *J. Chem. Phys.*, 109(19):8218, 1998.

- [206] C. M. Suomivuori, N. O. Winter, C. Hättig, D. Sundholm, and V. R. Kaila. Exploring the Light-Capturing Properties of Photosynthetic Chlorophyll Clusters Using Large-Scale Correlated Calculations. *J. Chem. Theory Comput.*, 12(6):2644–2651, 2016.
- [207] A. Szabo and N. S. Ostlund. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*. Dover Publ., Mineola, New York, 1 edition, 1996.
- [208] M. Takemoto, H. E. Kato, M. Koyama, J. Ito, M. Kamiya, S. Hayashi, A. D. Maturana, K. Deisseroth, R. Ishitani, and O. Nureki. Molecular Dynamics of Channelrhodopsin at the Early Stages of Channel Opening. *Plos One*, 10(6):e0131094, 2015.
- [209] Y. Tawada, T. Tsuneda, S. Yanagisawa, T. Yanai, and K. Hirao. A Long-Range-Corrected Time-Dependent Density Functional Theory. *J. Chem. Phys.*, 120(18):8425–8433, 2004.
- [210] E. Thyrgaug, K. Žídek, J. Dostál, D. Bína, and D. Zigmantas. Exciton Structure and Energy Transfer in the Fenna-Matthews-Olson Complex. *J. Phys. Chem. Lett.*, 7(9):1653–1660, 2016.
- [211] Y. Tong, C. G. Ji, Y. Mei, and J. Z. H. Zhang. Simulation of NMR Data Reveals That Proteins Local Structures Are Stabilized by Electronic Polarization. *J. Am. Chem. Soc.*, 131:8636–8641, 2009.
- [212] D. E. Tronrud, M. F. Schmid, and B. W. Matthews. Structure and X-ray Amino Acid Sequence of a Bacteriochlorophyll a Protein from *Prosthecochloris Aestuarii* Refined at 1.9 Å Resolution. *J. Mol. Biol.*, 188(3):443–454, 1986.
- [213] M. Umetsu, Z. Wang, K. Yoza, M. Kobayashi, and T. Nozawa. Interaction of Photosynthetic Pigments with various Organic Solvents 2. Application of Magnetic Circular Dichroism to Bacteriochlorophyll a and Light-Harvesting Complex 1. *Biochim. Biophys. Acta*, 1457(3):106–117, 2000.
- [214] O. Valsson and C. Filippi. Photoisomerization of Model Retinal Chromophores: Insight from Quantum Monte Carlo and Multiconfigurational Perturbation Theory. *J. Chem. Theory Comput.*, 6:1275–1292, 2010.
- [215] O. Valsson, C. Filippi, and M. E. Casida. Regarding the Use and Misuse of Retinal Protonated Schiff Base Photochemistry as a Test Case for Time-Dependent Density-Functional Theory. *J. Chem. Phys.*, 142(14), 2015.
- [216] M. K. Verhoeven, C. Bamann, R. Blöcher, U. Förster, E. Bamberg, and J. Wachtveitl. The Photocycle of Channelrhodopsin-2: Ultrafast Reaction Dynamics and Subsequent Reaction Steps. *ChemPhysChem*, 11(14):3113–3122, 2010.

- 
- [217] O. Volkov, K. Kovalev, V. Polovinkin, V. Borshchevskiy, C. Bamann, R. Astashkin, E. Marin, A. Popov, T. Balandin, D. Willbold, G. Büldt, E. Bamberg, and V. Gordeliy. Structural Insights Into Ion Conduction by Channelrhodopsin 2. *Science*, 358(6366), 2017.
- [218] E. Walczak, B. Szefczyk, and T. Andruniów. Geometries and Vertical Excitation Energies in Retinal Analogues Resolved at the CASPT2 Level of Theory: Critical Assessment of the Performance of CASSCF, CC2, and DFT Methods. *J. Chem. Theory Comput.*, 9(11):4915–4927, 2013.
- [219] M. Wanko, M. Hoffmann, J. Frähmcke, T. Frauenheim, and M. Elstner. Effect of polarization on the opsin shift in rhodopsins. 2. empirical polarization models for proteins. *J. Phys. Chem. B*, 112:11468–11478, 2008.
- [220] M. Wanko, M. Hoffmann, T. Frauenheim, and M. Elstner. Computational photochemistry of retinal proteins. *J. Comput. Aided Mol Des.*, 20(7):511–518, 2006.
- [221] M. Wanko, M. Hoffmann, T. Frauenheim, and M. Elstner. Effect of polarization on the opsin shift in rhodopsins. 1. a combined QM/QM/MM model for bacteriorhodopsin and pharaonis sensory rhodopsin II. *J. Phys. Chem. B*, 112:11462–11467, 2008.
- [222] M. Wanko, M. Hoffmann, P. Strodel, A. Koslowski, W. Thiel, F. Neese, T. Frauenheim, and M. Elstner. Calculating Absorption Shifts for Retinal Proteins: Computational Challenges. *J. Phys. Chem. B*, 109(8):3606–3615, 2005.
- [223] A. Warshel and M. Levitt. Theoretical Studies of Enzymic Reactions: Dielectric, Electrostatic and Steric Stabilization of the Carbonium Ion in the Reaction of Lysozyme. *J. Mol. Biol.*, 103:227–249, 1976.
- [224] H. C. Watanabe, K. Welke, F. Schneider, S. Tsunoda, F. Zhang, K. Deisseroth, P. Hegemann, and M. Elstner. Structural Model of Channelrhodopsin. *J. Biol. Chem.*, 287(10):7456–7466, 2012.
- [225] H. C. Watanabe, K. Welke, D. J. Sindhikara, P. Hegemann, and M. Elstner. Towards an Understanding of Channelrhodopsin Function: Simulations lead to Novel Insights of the Channel Mechanism. *J. Mol. Biol.*, 425(10):1795–814, 2013.
- [226] W. Weber and W. Thiel. Orthogonalization Corrections for Semiempirical Methods. *Theo. Chem. Acc.*, 103(6):495–506, 2000.
- [227] K. Welke, J. S. Frähmcke, H. C. Watanabe, P. Hegemann, and M. Elstner. Color Tuning in Binding Pocket Models of the Chlamydomonas-Type Channelrhodopsins. *J. Phys. Chem. B*, 115(50):15119–15128, 2011.

- [228] K. Welke, H. C. Watanabe, T. Wolter, M. Gaus, and M. Elstner. QM/MM Simulations of Vibrational Spectra of Bacteriorhodopsin and Channelrhodopsin-2. *Phys. Chem. Chem. Phys.*, 15(18):6651, 2013.
- [229] N. O. C. Winter, N. K. Graf, S. Leutwyler, and C. Hättig. Benchmarks for 0–0 Transitions of Aromatic Organic Molecules: DFT/B3LYP, ADC(2), CC2, SOS-CC2 and SCS-CC2 Compared to High-Resolution Gas-Phase Data. *Phys. Chem. Chem. Phys.*, 15(18):6623–6630, 2013.
- [230] N. O. C. Winter and C. Hättig. Scaled opposite-spin cc2 for ground and excited states with fourth order scaling computational costs. *J. Chem. Phys.*, 134(18):184101, 2011.
- [231] D. Wöhrle, M. W. Tausch, and W.-D. Stohrer. *Photochemie: Konzepte, Methoden, Experimente*. John Wiley & Sons, Weinheim, 2 edition, 1998.
- [232] F.-E. C. Wolff. *Computer Simulations of Light-Triggered Processes of Chromophores in Complex Environments*. PhD thesis, Karlsruhe Institute of Technology (KIT), Karlsruhe, 2018.
- [233] T. Wolter, M. Elstner, S. Fischer, J. C. Smith, and A.-N. Bondar. Mechanism by which Untwisting of Retinal Leads to Productive. *J. Phys. Chem. B*, 119:2229–2240, 2015.
- [234] T. Wolter, K. Welke, P. Phatak, A. N. Bondar, and M. Elstner. Excitation Energies of a Water-Bridged Twisted Retinal Structure in the Bacteriorhodopsin Proton Pump: A Theoretical Investigation. *Phys. Chem. Chem. Phys.*, 15(30):12582–12590, 2013.
- [235] B. Yan, J. L. Spudich, P. Mazur, S. Vunnam, F. Derguini, and K. Nakanishi. Spectral Tuning in Bacteriorhodopsin in the Absence of Counterion and Coplanarization Effects. *J. Biol. Chem.*, 270(50):29668–29670, 1995.
- [236] T. Yanai, D. P. Tew, and N. C. Handy. A new Hybrid Exchange-Correlation Functional using the Coulomb-Attenuating Method (CAM-B3LYP). *Chem. Phys. Lett.*, 393(1-3):51–57, 2004.
- [237] H. Yu and Q. Cui. The Vibrational Spectra of Protonated Water Clusters: A Benchmark for Self-Consistent-Charge Density-Functional Tight Binding. *J. Chem. Phys.*, 127(23), 2007.
- [238] H. Yu and W. F. Van Gunsteren. Accounting for Polarization in Molecular Simulation. *Comput. Phys. Commun.*, 172(2):69–85, 2005.
- [239] M. C. Zerner, G. H. Loew, R. F. Kirchner, and U. T. Mueller-Westerhoff. An Intermediate Neglect of Differential Overlap Technique for Spectroscopy of Transition-Metal Complexes. Ferrocene. *J. Am. Chem. Soc.*, 102(2):589–599, 1980.

- [240] J.-p. Zhang, R. Fujii, P. Qian, T. Inaba, T. Mizoguchi, Y. Koyama, K. Onaka, Y. Watanabe, and H. Nagae. Mechanism of the Carotenoid-to-Bacteriochlorophyll Energy Transfer via the S1 State in the LH2 Complexes from Purple Bacteria. *J. Phys. Chem. B*, 104(15):3683–3691, 2000.