Computer Simulations of Light-Triggered Processes of Chromophores in Complex Environments

Zur Erlangung des akademischen Grades einer
DOKTORIN DER NATURWISSENSCHAFTEN

(Dr. rer. nat.)

Karlsruhe Institute of Technology

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

genehmigte

DISSERTATION

von

Franziska-Elisabeth Charlotte Wolff
aus
Bretten

Referent: Prof. Marcus Elstner
Korreferent: Prof. Hans-Achim Wagenknecht
Tag der mündlichen Prüfung: 16.07.2018

Copyright

This work was a cooperation with Dr. Yanan Guo, where both authors contributed equally. The work was shared as follows: Franziska E. Beyle (Wolff) conducted the simulations and part of the analysis of ChR2 WT, while Yanan Guo performed the simulations and analysis of ChR2-C128T mutant and part of the analysis of ChR2 WT.

Zusammenfassung

Zur Erforschung von biologischen Prozessen und Entwicklungen von neuen Applikationen, ist das Verständniss von photophysikalischen und photochemischen Prozessen von Licht absorbierenden Molekülen (Chromophoren) in komplexen Umgebungen essentiell. Das maßgebliche Ziel dieser Arbeit ist das tiefere Verständnis von Absorptions, Fluoreszenz und Isomerisierungs Prozessen in unterschiedlichen Umgebungen und den Umgebungseinfluss auf die strukturellen Eigenschaften des Chromophors aufzudecken. Dabei wird ein weites Spektrum an theoretischen Rechenmethoden eingesetzt. Aufwendige und genaue *ab initio* Methoden werden für die korrekte Beschreibung der elektronischen Struktur des angeregten Zustandes verwendet. Wohingegen rechnerisch weniger anspruchsvolle, semi-empirische Methoden ein dynamisches Bild des Chromophors bereitstellen und die Berechnung größerer System mit bis zu 150 Atomen ermöglichen.

Kanalrhodopsin-2 mit Retinal als Chromophor ist das am meist genutzte optogenetische Werkzeug der neuronalen Forschung. Seine Funktion ermöglicht die Licht gesteuerte Depolarisation von neuronalen Membranen. Jedoch ist dieser Mechanismus bis heute nicht komplett aufgeklärt und die weitere Erforschung erfordert eine genaue and detaillierte Struktur des Dunkel-Zustandes. Deshalb wurden im Rahmen dieser Arbeit, der Einfluss der Bindungstasche und der Einfluss der strukturellen Konformation der Bindungstasche auf das Absorptionsverhalten untersucht. Es konnten mehrere spezifische strukturelle Merkmale und eine Temperaturabhängigkeit des Kanalrhodopsin-2 festgestellt werden.

Im Gegensatz zu Kanalrhodopsin-2, ist das Histidin Kinase Rhodopsin-1 (HKR1) fähig zwei stabile Zustände innerhalb des Photozyklus auszubilden. Dies ermöglicht ein Umschalten zwischen diesen zwei stabilen Zuständen des Proteins, wodurch es ein vielversprechendes Protein für die Entwicklung weitere Applikationen für die Optogenetik ist. Auch hier sind die Proteinstruktur und der Funktionsmechanismus noch nicht aufgeklärt. Aus diesem Grund wurde ein Homologiemodell mit einer verlässlichen Retinalstruktur und Bindungstasche gebaut und durch den Vergleich der Anregungsenergie mit der Referenz Bakteriorhodopsin verifiziert. Somit konnte die all-*trans* zu 13-*cis* Isomerisierung des Retinals mit *ab initio* Methoden untersucht werden. Infolgedessen konnte eine 13-*cis* Struktur des Retinals erhalten werden, die eine weitere Untersuchung des einzigartigen zweiten stabilen Zustandes des Fotozyklus erst ermöglicht.

Da *ab initio* Methoden zu rechenaufwendig für die Simulation des dynamischem Verhaltens von Systemen wie Retinal sind, wurde die semi-empirische OM2/MRCI Methode, die eine dynamische Beschreibung im angeregten Zustand ermöglicht, getestet. Sowie für die korrekte Bestimmung des Isomerisierungsmechanismus und der Lebenszeit im angeregten Zustand eines pronotierten Schiffbase Modells verifiziert.

Im letzten Teil dieser Arbeit wurde der solvatochrome Effekt des fluoreszenten Moleküls Flugi-2 analysiert. Die dynamische Beschreibung von Chromophoren im angeregten Zustand mit der gleichzeitigen eindeutigen Beschreibung des Lösungsmittel war bis jetzt nicht möglich und wurde erstmals in dieser Arbeit durchgeführt. Dazu wurden klassische Methoden verwendet um Geometrien im angeregten Zustand zu erhalten, wobei die elektronische Struktur mit semi-empirischen Methoden berechnet wurde. Dieses Konzept ermöglicht die Untersuchung des strukturellen Einflusses des Lösungsmittels auf den Chromophor Flugi-2. Ferner konnte gezeigt werden, dass implizite Solventmodelle die Wechselwirkung zwischen Lösungsmittel und Chromophore nicht korrekt beschreiben können. Bis zukünftige Methoden eine dynamische Beschreibung des Chromophors mit der expliziten Beschreibung der Umgebung ermöglichen, sind auf klassischen Methoden basierte Ansätze, wie in dieser Arbeit gezeigt, sinnvoll um die strukturellen Zusammenhänge zwischen Lösungsmittel und Chromophor zu beschreiben. In dieser Arbeit sind Absorptions-, Fluoreszenz- und Isomerisierungsprozesse in verschiedenen komplexen Umgebungen mit Hilfe unterschiedlicher theoretischer Methoden erfolgreich erforscht worden.

Abstract

The understanding of photophysical and photochemical processes of light absorbing molecules (chromophores) in complex environments is crucial for the development of new research tools for biological processes. The goal of this thesis is to gain a deeper knowledge of absorption, fluorescence and isomerization processes in different environments and the influence of the environment to structural properties of the chromophore. This is achieved by the application of a wide spectrum of theoretical computational methods. High level *ab initio* methods are required for the correct description of the excited state electronic structure. The computational less costly semi-empirical methods provide a dynamic picture of the chromophores and enables the enlargement of the considered systems.

The mostly used chromophore as optogenetic tools for the neuroscience research is retinal, which is found in rhodopsins. Channelrhodopsin-2 (ChR-2) is considered as the most used rhodopsin in this field. Its function enables the light triggered depolarization of neuronal membranes. However, the mechanism is still not clear and further investigation of the photocycle requires an accurate and detailed description of the dark-state structure. Therefore, the influence of the binding pocket of the retinal in ChR-2 and the effect of the structural fine tuning of the binding pocket on the absorption has been investigated. As a result, several structural characteristics of the protein and its temperature dependency have been determined.

In contrast to ChR-2, the histidin kinase rhodopsin-1 (HKR1) is able to form two stable states during the photocycle. This enables bimodal switching between the two stable states of the protein and so HKR1 is a promising protein for the development of new optogenetic tools. In HKR1, the structure and function mechanism is poorly understood. Thus, a homology model with a reliable retinal structure and its binding pocket has been built and successfully validated by comparison of the absorption to the absorption of bacteriorhodopsin as reference. This structure allows the investigation of the the all-*trans* to 13-*cis* isomerization of the retinal by using high level *ab initio* methods. This detected 13-*cis* retinal structure enables further investigations of the unique second stable state in the HKR1 photocycle.

The high level *ab initio* methods are computationally too demanding for the simulation of

the dynamic behavior of the isomerization of systems like retinal embedded in rhodopsins. Therefore, the faster semi-empirical OM2/MRCI method, has been tested and benchmarked for the determination of the isomerization mechanism and the excited state life-time of a protonated Schiff base model.

In the last part of this work, the solvatochromic effect of the fluorescent molecule Flugi-2 has been investigated. The dynamical treatment of chromophores in the excited state, while considering the explicit surrounding is not possible to this date. Thus, new methods first applied the classical methods have been used to provide geometries in the excited state in combination with the electronic structure calculated by semi-empirical methods. Hence, resulting in the investigation of the structural influence of the DMSO solvent on the Flugi-2 chromophore. It has been demonstrated, that implicit solvent models fail to describe the correct interaction between solvent and chromophore. Until future methods enable a dynamical treatment of the chromophore with an explicit description of the environment, classical methods as described above are a reasonable way to sample the structural correlation between solvent and chromophore. In this work, it has been successfully investigated absorption, fluorescence and isomerization processes of chromophores in different complex environments by using a wide range of computationally methods.

Contents

Part I

Introduction

1 | Light-Triggered Processes - Interaction of Light with Matter

The interaction of light with matter is the interaction of an electromagnetic wave with atoms and molecules and leads to reflection, refraction or absorption. During the reflection and refraction the energy of the light is not absorbed by the matter, whereas in the absorption process the light interacts directly with the matter and opens new dimensions of photophysical and photochemical processes.

These processes are sensitive to the environment and provide a direct indication of the surrounding of an absorbing molecule, called chromophore. The use of this property opens new opportunities for the development of research tools for biological processes. However, extended knowledge of these interactions is necessary and part of this research.

In this work, the focus lies on the interaction of light in the UV/vis region of the electromagnetic spectrum, which includes the near-ultraviolet (200 nm-400 nm) and the visible regions (400 nm-700 nm) of the spectrum.

Due to the energy range of the electromagnetic wave, the interaction between matter and light leads to electronic and vibrational excitations. Only the absorption of light leads to the photophysical and photochemical processes. Here, the electric component of the electromagnetic wave interacts with the electronic structure of the matter, therefore the interaction and the subsequent processes have to be explained and described by quantum mechanics. Since absorption is a quantum effect, only one photon can be absorbed by a single molecule.[1–4]

1.1 Jabłoński Diagram and Photophysical Processes

An overview of photophysical light-triggered processes is given by the so-called Jabłoński diagram. Here, the electronic transition of the valence electrons in the higher energetic electronic states S_n (thick horizontal lines) and their radiation or radiation-less decay back to the ground state S_0 are described. Radiation induced transitions are called *absorption* (red arrows), *emission* (green arrows) or *phosphorescence* (blue arrow), while radiation-less

decay processes are represented as wavy arrows.

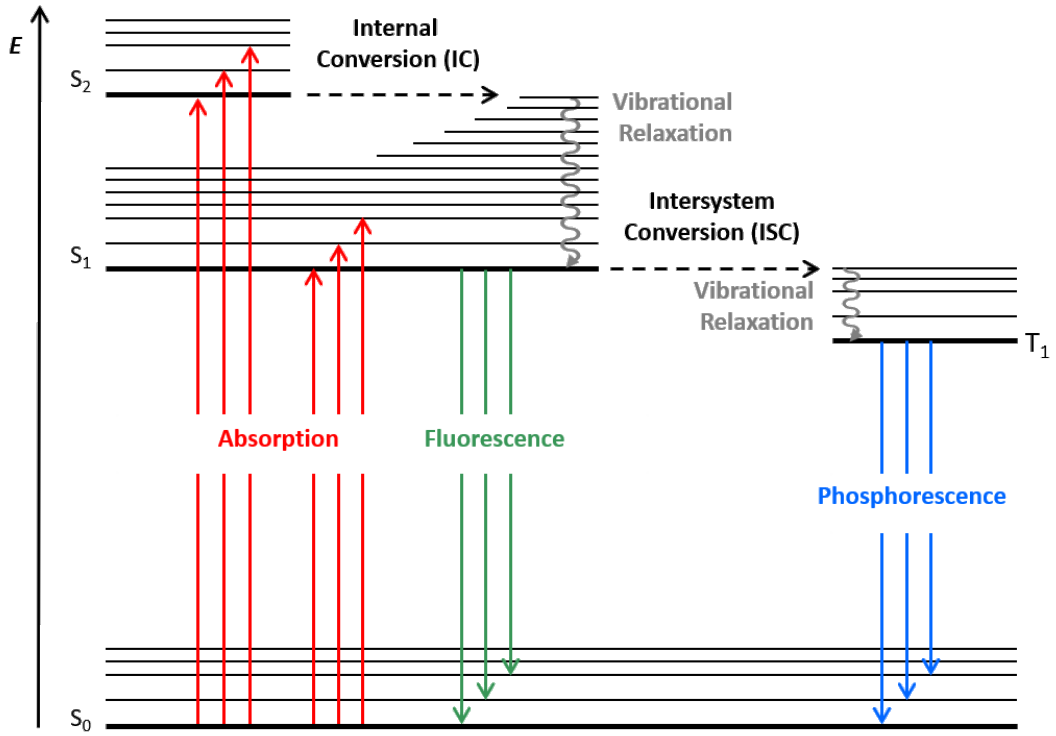


Figure 1.1.1: Jabłoński diagram: State diagram summarizing photophysical and photochemical processes. The minima of the particular electronic states (thick horizontal lines), together with vibrational energy levels (thin horizontal lines) are depicted. Furthermore, in the vertical lines the radiative and radiation-less transitions occurring between the electronic states are shown.

The processes are initiated by the absorption of light (red arrows), which transfers the system from the ground state S_0 to higher electronic states S_n. The absorption is precisely explained in the following section. The thin horizontal lines represent the vibrational state (ν_i) of each electronic state S_n. The absorption transition can also happen in higher lying vibrational state of the excited singlet state, than it will be called vibronic transitions. A following relaxation of the system via a *vibrational relaxation*, which transfers the energy to the environment, leads to the lowest vibrational state (ν_0) of the particular electronic state. Vibrational relaxation is the energetic transition of vibrational energy of the system to the surrounding and can occur in any electronic state. The absorption occur only from the ground state to a singlet excited state, while the excitation by light (UV/vis) within the same electronic state is forbidden and is only observed in the near-infrared region of the electromagnetic spectrum. The excitation produces no large significant structural changes, only the bond lengths and angles of the atomic bonds vary in the excited state.

Starting from the singlet excited state the system can relax by a so-called *internal conversion* to lower electronic singlet states. This relaxation is a fast process in a time scale of femtoseconds to nanoseconds (see Table 1.1.1), since it is an isoenergetic radiation-less transition between two states of the same multiplicity. After an internal conversion the system

is in a higher vibrational state of the lower electronic singlet state, i.e. internal conversion happens from $S_2(\nu_0) \rightarrow S_1(\nu_{i>0})$. Via vibrational relaxation the system transfers energy in form of vibrational energy or heat to the environment and relaxes to the lowest vibrational state of the electronic state.

From the lowest singlet excited state the system can relax back to the singlet ground state via radiation-less decay, *internal conversion* with a following vibrational relaxation or relaxation via a radiative transition, which is called fluorescence. The emission of the photon occurs only from the lowest excited singlet state (S_1) to the singlet ground state (S_0) (Kasha's rule). This process is a slow process with a time scale of nanoseconds. A more detailed description of this process is in the following section.

An unusual process is the change of the system multiplicity during a transition, which is called intersystem crossing. It is an isoenergetic radiation-less transition with a spin flip leading to a state with different multiplicity. This is an irreversible process, because of the increasing entropy. Intersystem crossing can occur from singlet to triplet states and vice versa ($T \rightarrow S$ or $S \rightarrow T$). In fact, these processes are spectroscopically forbidden, but due to the magnetic coupling of the electron spin with the angular momentum, i.e. the spin-orbit coupling, this process can occur. Hereby, the rule of El-Sayed states, that the transition in a state with different multiplicity contains of the same time a transition to an orbital with different symmetry (i.e. $\pi \rightarrow n^*$ or $n \rightarrow \pi^*$)

The phosphorescence is a radiative transition from the lowest triplet state T_1 to the singlet ground state. Due to the change of the spin multiplicity this is the slowest photophysical process with a time scale from nanoseconds to seconds.[5]

Table 1.1.1: Time scales of photophysical processes

Process	Time Scale [s]
Absorption	10^{-15}
Internal Conversion	10^{-12}-10^{-6}
Intersystem Crossing ($S \rightarrow T$)	10^{-12}-10^{-6}
Intersystem Crossing ($T \rightarrow S$)	10^{-9}-10^{1}
Vibrational Relaxation	10^{-13}-10^{-12}
Fluorescence	10^{-9}-10^{-7}
Phosphoresence	10^{-6}-10^{-3}

1.1.1 Absorption and Fluorescence

In the physical description of the absorption process of light by matter, the valence electrons of a molecule or the electronic band structure of matter interact with the electronic component

of the light wave. The absorbed photon excites the valence electrons of a molecule, which has to have a quantum transition with exact the same energy as the photon. Therefore, absorption is a frequency dependent process. This means, that the electronic vector of the light wave with the frequency ν interacts with the charged electrons of the matter.

In organic molecules, the absorption takes place in the valence electrons mainly of the 2s and 2p orbitals of the carbon atoms, rarely from the valence electrons of the oxygen or nitrogen atoms. The concept in organci molecules can be explained by the molecular electronic transition. Here, the transition leads to an excitation of an electron from the highest occupied orbital (HOMO) to the lowest unoccupied orbital (LUMO). Three types of transitions can be considered: i) Transition between n, σ and π type orbitals, ii) charge transfer transition and iii) transitions to *d* or *f* orbitals of the atoms.

The colour of materials arises, when only specific wave lengths are absorbed by the matter, while the remaining light is reflected. For example a molecule appears red, when all wave lengths of light are absorbed by the molecule expect the red light. When shining on the material with white light and the material absorbs at, for example 480 nm (blue light), the colour perceived is the complementary colour, yellow in this case.

Fluorescence is a similar process, but here a photon is emitted by a spontaneous process after a previous absorption of light with a simultaneously back transition of the system in the electronic ground state (see Figure 1.1.1). Here, the energy of the photon equals the energy gap between the electronic state S_1 and S_0. Since the transition takes place between two singlet states only and is therefore spectroscopically allowed, the time scale is in the order of ns (see Table 1.1.1). However, the system stays in the excited electronic state, while it relaxes from a higher vibrational state to the lowest vibrational state of the excited electronic state before it can emit the photon. The time, in which the system remains in the excited state after absorption is called the fluorescence life time and is molecule and matter dependent. Furthermore, the life time can be theoretically predicted with "Fermi's Golden rule". It determines the rate of the transition between two states by applying a perturbation to the initial state.

The relaxation process leads to energy loss, therefore the fluorescence wave length is always longer (lower in energy) than the absorption wave length. This consequential difference between the absorption wave length and the fluorescence wave length is called Stokes shift and is also specific to the molecule or material.

Absorption and Fluorescence are both adiabatic processes, since the transition occurs between two adiabatic states.

1.1.2 Solvatochromism

One astonishing interaction between light and matter the solvatochromism. It is the solvent dependent shift of the absorption or emission wave length. The energy shift is dependent on the polarity of the solvent surrounding the chromophore. The effect is mainly the result of electrostatic interactions and therefore occurs only in chromophores with different dipole moments between the ground and excited state. Solvent iwth different polarity leads to a different ground or excited state stabilization of the chromophore, based on the dipole moment of the particular state. More in detail, a solvent with an higher polarity interacts stronger with the chromophore and stabilizes the particular state, which leads to a blue shift of the transition wave length. Contrarily, a non-polar solvent decreases the stability of this state and leads to an read shift. This can lead to an astonishing visual effect when a chromphore is solved in different solvents, which leads to a variety of solution colours. The specific interaction between solvent and chromophore can be measured by the absorption spectrum.

In the fluorescence process the solvatochromism effect is bigger the larger the change of the dipole moment is. This results in the reorganisation energy of the solvent to the excited state. Based on the Franck-Condon principle the chromophore in the excited state has an increased dipole moment, but the surrounding solvent is not in the equilibrium with the excited chromophore and a reorganisation of the solvent to the new dipole moment of the chromophore leads to a decrease of energy (see Figure 1.1.2).

Chromophores with a charge transfer transition show the largest solvent sensitivity, since excitation of this type of chromophores leads to relocation of formal charges over the chromophore.

1.1.3 Franck-Condon Principle and Transition Moments

The Franck-Condon principle is a quantum mechanical principle and allows the explanation of the transition probability between two electronic states. The principle is illustrated in Figure 1.1.3.

Basic assumption for the Franck-Condon principle is the Born-Oppenheimer approximation, which is based on the observation, that the movement of the electrons is orders of magnitude faster compared to the movement of the nuclei. This indicates, that the positions of nuclei do not change during electronic transitions.

The probability of the electronic transition is dependent on the shape of the wave function, a transition is rather probable when the overlap of the wave functions of the states is big. This becomes clearer, when the origin of the electronic transition is considered.

As the electromagnetic component of the light wave interacts with the electrons and the

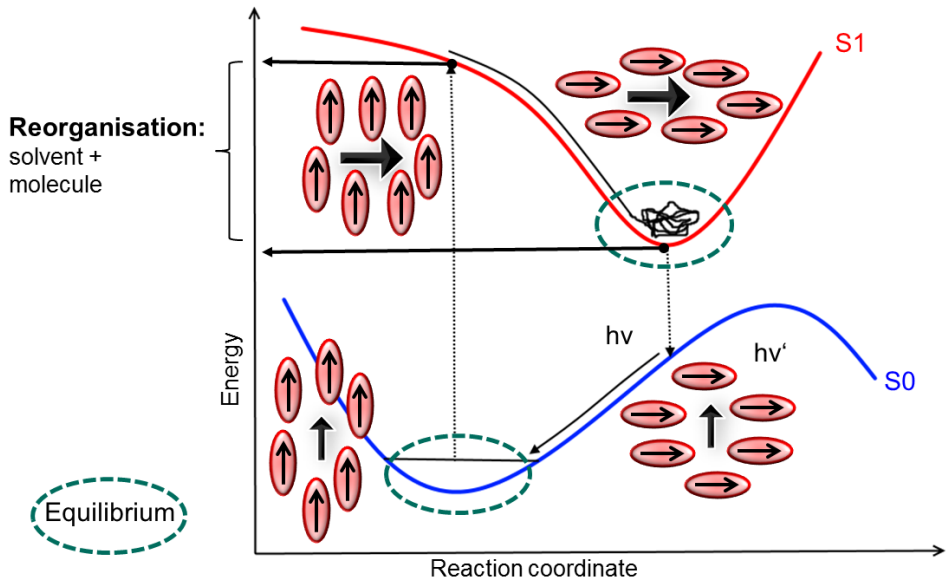


Figure 1.1.2: Scheme of the events, when a chromophore in the excited state relaxes and the solvent "reacts" to the excited chromophore.

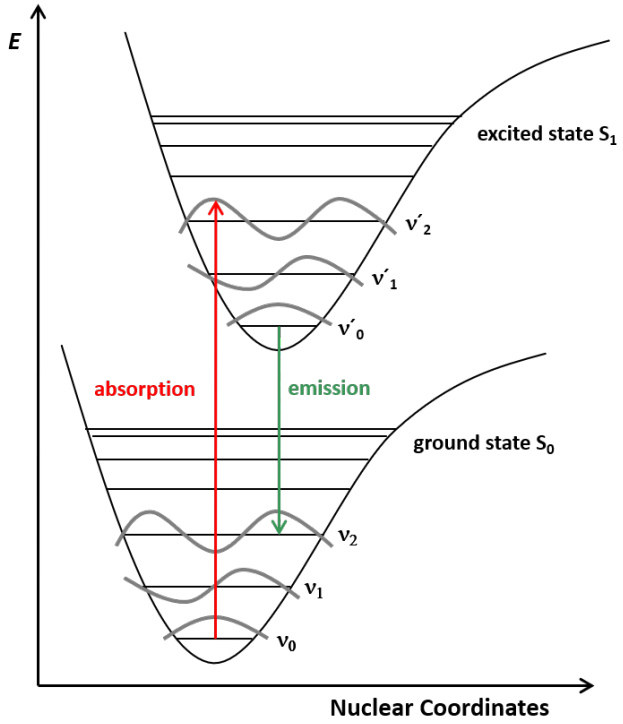


Figure 1.1.3: Franck-Condon principle: The electrons are much heavier than the nuclei, therefore in the Franck-Condon principle it is assumed, that the transition process is so fast, that it occurs in the electrons, while the nuclei stay at their positions. The nuclei "react" to the new electronic structure and relaxing to their new minima.

nuclei, the transition can be calculated based on the dipole moment operator $\widehat{M}$ of the system. The dipole moment operator contains the charges of the nuclei (Z_k) and the positions of the electrons (r_i) and nuclei (r$_k$). Therefore, the transition moment ($M_{n,0\rightarrow m,\nu'}$) can be computed from the expectation value of the dipole moment operator between two different electronic states (n and m) (see Equation 1.1). In this connection, the transition moment represents the coupling of the ground state wave function with the excited state wave function.

$$\begin{aligned} M_{n,0\rightarrow m,\nu'} &= \left\langle \Psi_{n,0}|\widehat{M}|\Psi_{m,\nu'} \right\rangle \\ \widehat{M} &= e\left(\sum r_i - \sum Z_k r_k\right) \end{aligned} \tag{1.1}$$

Due to the Born-Oppenheimer approximation the electronic ($\Psi_e l$) and the nuclear wave function (ξ) can be separated, since the nuclear wave function does not depend on the positions of the electrons. Thereby, Equation 1.1 is simplified to Equation 1.2:

$$M_{n,0\rightarrow m,\nu'} = e\left\langle \Psi_{el,n}|\sum r_i|\Psi_{el,m}\right\rangle \langle \xi 0|\xi_{\nu'}\rangle \tag{1.2}$$

Now, the transition probability is the product of the electronic transition moment ($e\left\langle \Psi_{el,n}|\sum r_i|\Psi_{el,m}\right\rangle$) and the nuclear overlap integral ($\langle \xi 0|\xi_{\nu'}\rangle$). The vibrational transition given by the overlap integral of the nuclear wave functions has to be unequal to zero, while otherwise the transition moment vanishes and the transition probability is zero. The value of the overlap integral increases along with the similarity of the nuclear vibrational states.

In Figure 1.1.3 another characteristic of the electronic transition is illustrated. The energetic minimum position of the nuclei in the electronic excited state is different from the positions in the ground state and this leads to the relaxation of the nuclei after the electronic excitations. The nuclei react to the new energetic situation given by the electrons in the excited state and they reach their new minimum configuration after a certain time, this is called vibrational relaxation.[1–4]

1.1.4 Challenges of Characterizing Absorption and Fluorescence Processes in Theory

In the theoretical description of absorption and fluorescence the zero point vibrational energy is generally neglected, because the excitation energy is determined on static structures. Furthermore, the vibrational excitations are not included in the calculations. However, absorption and also fluorescence are not static processes, rather the molecules fluctuate around their thermal equilibrium minimum. In order to resemble the absorption spectra, the absorption energy needs to be calculated on a set of geometries from this thermal equilibrium. But the *ab initio* quantum mechanics methods are computationally costly and

computation of this large number of excitation energies is only possible by semi-empirical methods. However, the semi-empirical methods can not be used as black box, but need to be benchmarked for their reliability on the applied system.

Furthermore, complications arise when photophysical and photochemical processes in proteins are characterized. The protein environment is crucial for the absorption of the chromophore and the followed photophysical and photochemical processes are strongly dependent on the surrounding amino acid composition. The interaction between the protein environment and the chromophore is crucial because the surrounding amino acids influence the geometry and the charge distribution of the chromophore and this needs to be considered exactly to enable a complete and exact picture of the absorption properties of the chromophore.

A general strategy is the treatment of the chromophore by quantum mechanical methods with a coincidental treatment of the environment by classical mechanic methods. However, still some effects are missing, i.e. the polarization of the charge distribution of the environment surrounding the chromophore. This leads to a systematically overestimation of the excitation energy.

Moreover, the amino acid side chain orientation can have a direct influence on the chromophore structure. The coupling between the side chain electronic structure (orbitals) and the chromophore electronic structure needs to be described explicitly, for example to describe hydrogen bonds between amino acids and the chromophore. Classical methods underestimate these hydrogen bonds as was seen in the studies of Bacteriorhodopsin (BR).[6] But the inclusion of amino acids in the QM zone results very quickly in system sizes of more than 100 atoms and this is not feasible with the high level quantum mechanics methods, so again a low level method is required. Therefore, semi-empirical methods need to be used with all their advantages and drawbacks.

Nevertheless, a deep knowledge about the utilized methods, their characteristics and drawbacks as well as a critical interpretation of the described system is crucial for successful research on photophysical and photochemical processes in proteins. A balance between accuracy of the method, choice of the zone described quantum mechanically and computation time is a challenging task.

In this work, several methods have been used to investigate absorption and fluorescence of a chromophore within complex environments. The aim of this research is to get a more detailed knowledge about the interactions between the environment and the chromophore and the resulting influence on the absorption and fluorescence behavior.[7]

1.2 Photochemical Process

The photochemical process is initiated from an excited state and results in a photoproduct that is chemically different from the educt. Relaxation of the system back to the ground state can occur via a radiation-less decay, a *Conical Intersection* or an *Avoided Crossing*. However, photochemical reaction paths are always in competition with photophysical processes and can only be efficient when they are energetically preferred and when they are faster than photophysical processes.

Due to the Born-Oppenheimer approximation and the adiabatic description of the processes, potential energy surfaces (PES) are obtained, which provide a more easily understandable picture. They describe the potential energy along a reaction coordinate and depict the concepts of photophysical and photochemical processes (see Figure 1.2.1) in a descriptive scheme.

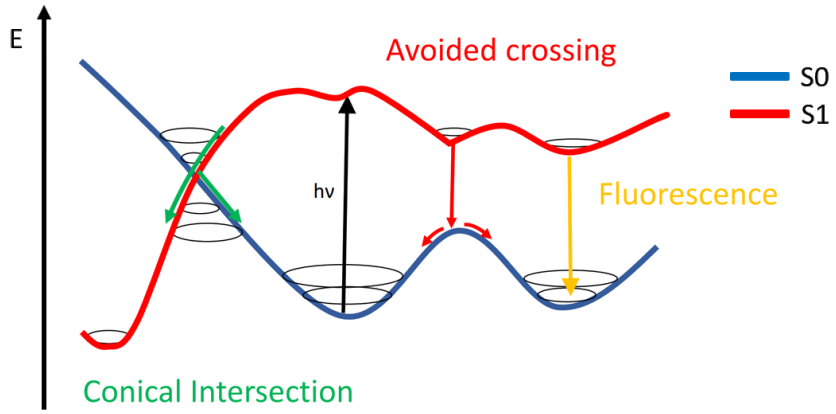


Figure 1.2.1: Photophysical and photochemical processes on the adaiabatic potential energy surface (PES) of a system. For more details see text.

For the radiation-less decay over a *Conical Intersection* the Born-Oppenheimer approximation breaks down and the adiabatic decription fails, since the wave function can not be described independently from the electrons. At the *Conical Intersection* small changes of the nuclear positions lead to changes of the electronic state. This is called non-adiabatic coupling and a non-adiabatic description is necessary to describe the transition.

Besides, the decay over a *Conical Intersection*, photochemical reactions can also relax over a thermal relaxation path back to the ground state, this is called *Avoided Crossing*. Usually, the system relaxes to a transition state in the ground state and leads to either the product or relaxes back to the educt.

An important parameter in the photochemistry is the quantum yield Φ, which determines the ratio of the number of photochemical events n_x occuring and the number n_P at the irradiation wave length λ.

$$\Phi(\lambda) = n_x/n_P \tag{2.1}$$

The task of theoretical methods is to predict these PES in order to gain insight about the photochemical processes, the reaction mechanisms and their time constants. Therefore, high level *ab initio* methods are necessary to ensure a correct description of the PES, however semi-empirical methods provide extended chance to characterize the photochemistry in complex environments.[1,3]

2 | The Photoreceptors of *Chlamydomonas Reinhardtii*

The first step of the vision process is the absorption of one photon by the chromophore in a photo-receptor leading to a photochemical reaction of the chromophore. This impressive physicochemical process combines two optical processes, the absorption and the photochemical reaction. The isomerization of the chromophore initiates a photocycle, where several rearrangements happen within the protein. A cascade of biochemical reactions starts and thus the light signal is transferred to the cell. Such photocycles are the basis of diverse biological functions like the vision process in vertebrates and invertebrates, light-triggered growth in plants, phototaxis, light sensing and ion pumping and gating in eubacteria and archae.

The transformation from light energy to mechanical energy over the retinal on the molecular level is only possible because of the surrounding protein. In other words the optical process is only possible due to the complex environment around the retinal chromophore. It features the effective, fast and specific double bond isomerization in the retinal,[8] which is not self-evident since in solution all double bonds of retinal isomerize.

For the vision process, the rhodopsins are the fundamental protein, their part in the vision process is the direct interaction with the light. The chromophore in rhodopsins is always the retinal, which belongs to the molecule group of vitamin-A.[9] Retinal is covalently bound to lysine through a protonated Schiff base ($RSBH^+$) in all rhodopsins. The absorption property of the retinal molecule is determined by several structural factors: On the one hand, the planarity and the bond length alternation of the single and double bond in the retinal molecule determines the absorption wave length in protein environments. On the other hand, the hydrogen bonding interaction of the $RSBH^+$ is crucial for the absorption wave length. The electrostatic interaction between retinal and a protein environment with charged residues in a so-called binding pocket has an influence on the wave length as well. The rhodopsins occur only as membrane proteins, where the secondary structure is built by seven transmembrane helices permeating the membrane. Essential for the function of the rhodopsins is the proton transfer from the Schiff base to the protein environment.

Two types of rhodopsins are known, which show different structural aspects in the primary amino acid sequence, helical arrangement and the isomeric conformation of the retinal.

Type I rhodopsins can be found in prokaryotes like archae, bacteria and eukaryote microbes like fungi and algae and are therefore called microbial rhodopsins. Their biological function spans a wide field from ion pumping, phototaxis, ion channel gating up to photoreceptor functions for the photophobic reaction and enables the organisms to react to light.

Moreover, the type II rhodopsins have been found in higher eukaryotes and are called animal rhodopsins. They are responsible for the vision process in the organisms and acting as a G-protein, receptor kinase or are coupled to other signaling proteins.[10–12]

Among these two types of rhodopsins the composition of the amino acids in the binding pocket is similar, while the composition varies strongly between the two rhodopsin types. This is obvious, since the evolution of these two types took place in different evolutionary steps.

The retinal configuration in the dark-state of type I rhodopsin is present in the all-*trans* conformation. The isomerization to the 13-*cis* conformation (see Figure 2.0.1) in the initial step of the photocycle leads to the subsequent generation of intermediates. The Schiff base deprotonates stepwise and the protein reacts to the new retinal configuration with structural rearrangements, which leads at the end to the function mechanism of the rhodopsin. Finally, the protein relaxes back to the initial dark-state and the Schiff base reprotonates. Bacteriorhodopsin (BR) is the most studied microbial rhodopsin, its photocycle is well-known and is often used as reference for the investigation of the photocycle of other rhodopsins.

In type II rhodopsins, the retinal adopts a 11-cis conformation in the dark-state and light exposure leads to the all-*trans* conformation. The following steps are quite different from the microbial rhodopsin, since the retinal will be hydrolyzed from the apoprotein and this initiates the activation of a G-protein complex, which transfers the light stimulus into the cell.[13] Furthermore, the β ionone ring orientation differs between type I and type II rhodopsins. Type I rhodopsin has a 6-s-trans configuration around C7-C6 bond, type II rhodopsin has a 6-s-cis conformation.

The vision process in animals is initialized by the seven-transmembrane-helix rhodopsins, photoreceptors with covalently linked retinal. However, the signal transduction is different in vertebrates and invertebrates. In vertebrates the vision process takes place with a second messenger dependent signal transduction pathway via cyclic guanosine monophosphate (cGMP), which leads to the closure of the Na^{+}/Ca^{2+} cation channels and the hyperpolarization of the plasma membrane. In invertebrates the plasma membrane is depolarized over an activation of Phospholipase C and a subsequent release of diacyglycerol (DAG), which in turn activates the TRP/TRPL ion channels and leads to cation influx and the depolarization of the plasma membrane.

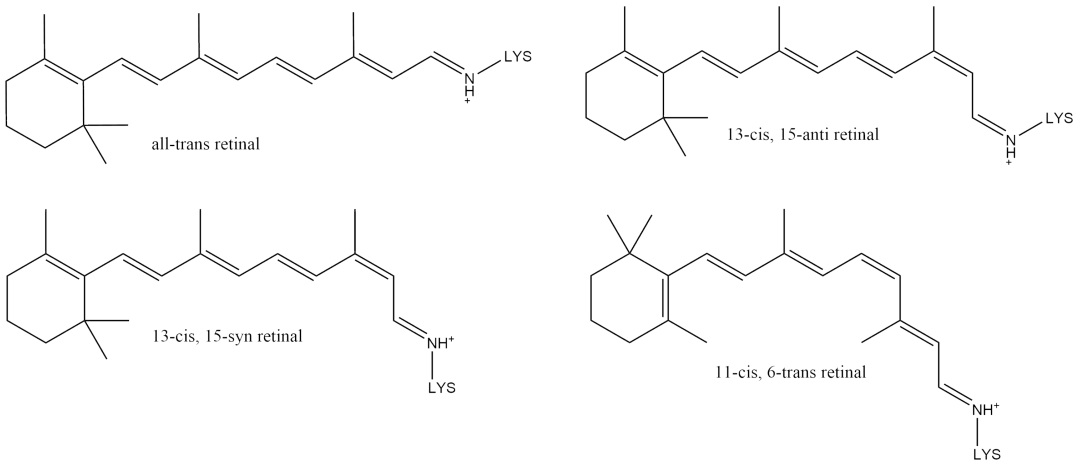


Figure 2.0.1: Different retinal configurations in type I and type II rhodopsins.

Detection of light is also important for microbial organisms. Not only is the rhodopsin structure different between animals (type II) and microbial organisms (type I), also the signal transfer mechanisms is based on different strategies. In microbial organisms the signal transduction is mediated by a so-called two-component system. This is a type of non-electrical signal transmission between sensor and regulator. This stimulus-response signaling mechanism is not only known from the "vision" process, but is a general strategy for signal transmission in single cell organisms. Usually it consists of a membrane bound histidin kinase domain, which senses the signal and a response regulator domain (RR) mediates the stimulus in the cell and activates the cell response. More in detail, the histidin kinase hydrolyzes a phosphoryl group from a bound ATP molecule to a histdin residue, while the RR domaine transfers this phosphoryl group to the response regulator's receiver domain, followed by a conformational change of the RR effector domaine, which triggers the cellular response to the signal.

It is supposed that the vision process in animal rhodopsins is evolved from the light-sensing system of ancestral unicellular flagellates, such as the algae *Chlamydomonas reinhardtii*. The vision process in algae could be the evolutionary connection between the animal and microbial vision strategies. *Chlamydomonas reinhardtii* has a unique visual system located in the "eye-spot" on the surface of the cell body. This organelle senses the light intensity and the incident angle during rotation of the algae and causes a motility reaction of the algae to the light.[14] This behavior is called phototaxis. In 1984, it had been determined, that the photoreceptor is rhodospin,[15] more precisely *Chlamydomonas reinhardtii* possess three different types of rhodopsins. In fact, one of the rhodopsin-types in the eye-spot acts as a Ca^{2+}-channel and depolarizes the plasma membrane by light exposure.[16,17] The higher concentration of Ca^{2+} in the cell is transferred to the flagella and induces a switch of flagella motion, which results in a change of directed movement of the cell. These rhodopsins called channelrhodopsins are related to microbial rhodopsins. Their biophysical properties

is described in more detail in the next section 2.1. The second type of rhodopsins are animal-like rhodopsins. They are coupled to a G-protein activation domain and are not part of the primary photoreceptor, but supposed to be involved in regulation of photosystem-I.[18,19] These types of rhodopsins are not part of this study. The third group of rhodopsins are more related to SRI and SRII of halobacteria and are coupled to histidine kinase and a prokaryotic-like response regulator domain. Thus, this rhodopsin group is called histidine kinase rhodopsin (HKR) and are described in chapter 7.

Since the first detection of the channelrhodopsins a series of Channelrhodopsins have been identified in different green alga[20–24]

2.1 Channelrhodopsin - a Light-Gated Ion Channel

In *Chlamydomonas reinhardtii*, two gene sequences Cop 3 and Cop 4 coding the channelrhodopsin 1 (ChR-1) and the ChR-2 have been identified.[16,17] Both are light-gated ion channels, which depolarize rapidly the cell membrane without the involvement of any transducer or second messenger. ChR-1 and ChR-2 differ in the sensing light wavelength, in the sensing light intensity, in the ion selectivity of the channel and in the delay time after the initial light flash.[25] This different behavior has been explained by the different behavioral response function like photophobic response and phototaxis.

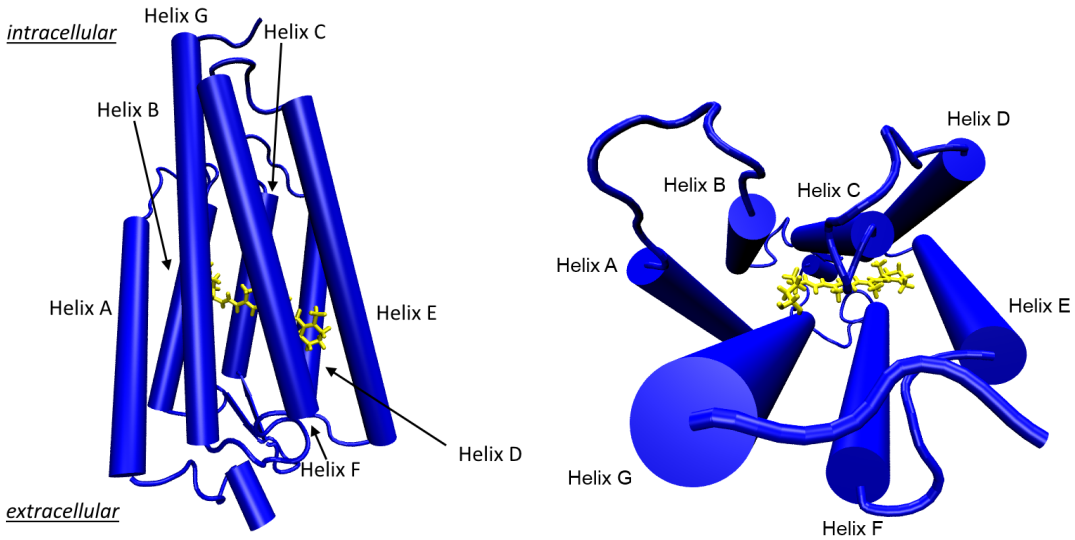
ChR-1 is characterized by faster kinetics, whereas ChR2 has a higher ion conductance. While, ChR-1 is a passive light gated proton channel,[16] ChR-2 is permeable for monovalent (Na^+, K^+) and divalent cations (Ca^{2+}). However, the photo currents of the channels are dependent on the atomic radius of the cations via an inverse relation.[17] The absorption wave length of ChR-1 has a maximum at 510 nm and is sensitive to light with high intensities, which implies the assignment for the photophobis response. ChR-2 is more sensitive to low light intensities with the absorption maximum at 470 nm and is responsible for the phototaxis of *Chlamydomonas reinhardtii*.[25]

Since ChR-2 shows remarkable expression levels in host cells and a pH independent absorption range between pH 4 and 9,[26] it has been applied to activate neurons under illumination and found a wide application as optogenetical tool.[27–29] Therefore, it is the best investigated channelrhodopsin.

2.1.1 ChR-2 Structure

Both, ChR-1 and ChR-2 are embedded as dimer in the membrane of the eye-spot, of which ChR-1 consists of 712 amino acids, where only one third of them form transmembrane helices (amino acids 76 to 309) while ChR2 has a slightly larger sequence (737 amino acids), but is otherwise very similar. At this point in time, the function of the long C-terminal

region remains unknown, but its truncation does not affect the photo current. Until late 2017 only a crystal structure of the chimera C1C2 was available, consisting of the first five helices of ChR1 and the last two helices of ChR2 available (PDB-Code: 3UG9[30]). Now, a crystal structure of ChR-2 has been published in late 2017,[31] but a structure for ChR-1 is still misssing. The chromophore in the channelrhodopsins is retinal, is covalently bound to a conserved lysine via a Schiff base (SB) and is mainly in an all-*trans* configuration, while the 13-*cis* isomer is the main photoproduct after irradiation. (see Figure 2.1.1)



(a) front view (b) top view

Figure 2.1.1: Overview of a channelrhodopsin monomer. The helical alignment is similar in all three channelrhodopsins; ChR1, ChR2, C1C2.

The opening of the cation channel is coupled to the isomerisation of the retinal and the following conformational changes in the protein.[32–36]

Several homology modeling studies[37] had been performed based on C1C2 crystal structure and compared to other microbial rhodopsins, in order to obtain structural information. The main difference between microbial rhodopsin structures (e.g. Bacterioshodopsin, bR) and the channelrhodopsin structure can be found in helices A and B, where channelrhodopsin has an untypical large number of glutamates (E82, E83, E90, E97, E101).
Some of these glutamates (E97, E101) and other polar residues (S52, Q56) build an electronegative pore between the helices A, B, C and G, which connects the extracellular region to the active site (see Figure 2.1.2). This pore has also been found in the ChR-2 crystal structure (see Figure 6.4.15a).

Mutations at this cavity affect the channel kinetics, but do not stop the photo current and therefore it is supposed to be part of the cation channel.[30,34,37–41] Also, Watanabe *et al.*[42] detected a cation biding site, which is located along the pore.

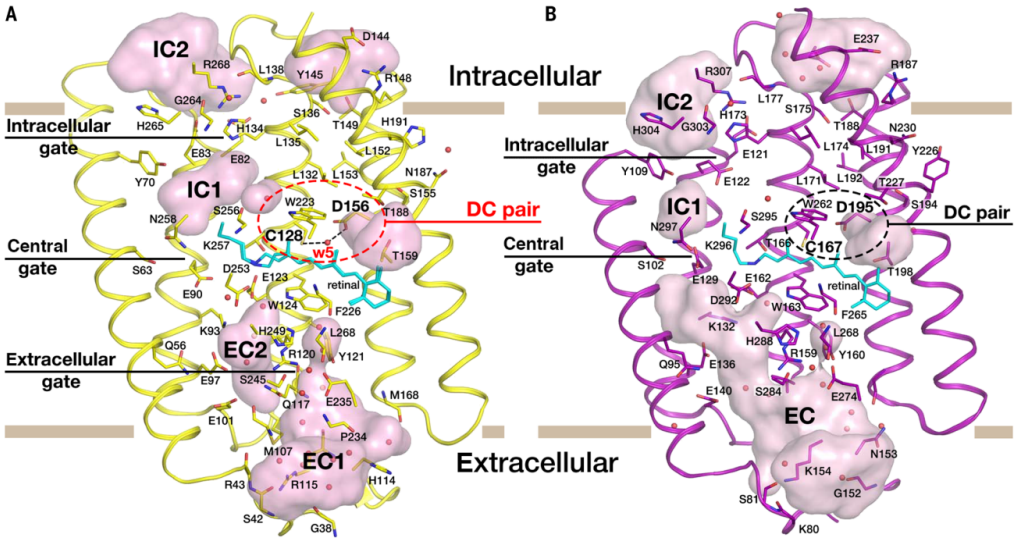


Figure 2.1.2: Comparison of the channelrhodopsin-2 (A) and the C1C2 chimera (B) crystal structure and the cavities of them. The picture is adapted from Volkov *et al.*[31]

Retinal and Active Site

The active site is formed by the retinal covalently linked to the protein via a $RSBH^+$ and the two counter ions, glutamate 123 (E123) and aspartate 253 (D253) (homologous residues in BR D85 and D212). Resonance Raman spectroscopy has revealed, that the $RSBH^+$ points towards the two counter ions in the dark-state.[43] In ChR2, E123 is probably deprotonated and therefore negatively charged, since mutation of E123 with a neutral residue results in a red shift of the absorption maximum,[17,26,30] while in ChR1 the same mutation has no effect.[44] In contrast, in C1C2 the counter ion E123 is neutral, based on empirical pK_a calculations[30] and spectroscopic measurements[43,45–47]

One important question, is the controversial debate about the hydrogen bond pattern of the $RSBH^+$.[43] The $RSBH^+$ could either be bound via a water molecule (as in bR) or form a direct salt bridge to one of the counter ions.[37,42,43,45–47] Resonance Raman-spectroscopic experiments suggest that the $RSBH^+$ is strongly hydrogen bonded to the counter ion.[43] Especially in QM/MM simulations of ChR2, a hydrogen bond from the $RSBH^+$ to a nearby water molecule was observed,[42] this is consistent with the down shifted Resonance Raman spectra of the $RSBH^+$ C=N vibration by H_2O/D_2O exchange.[43,48] This results from the increased water concentration in the vicinity of the $RSBH^+$ in QM/MM-calculations after several ns MD simulation as was reported in the C1C2 crystal structure.[42] In the active site in C1C2 crystal structure only one water molecule (w619)[30] at a distance about 4.4 Å is found. However, in ChR-2 crystal structure two water molecules (w1 and w2) are present near the $RSBH^+$ both within a distance of 4.5 Å from the $RSBH^+$.

The hydrogen bond pattern at the $RSBH^+$ affects the characteristics of the absorption

spectrum. The absorption spectrum of Channelrhodopsin shows three maxima at 479 nm, 445 nm and 415 nm, where the peak at 479 nm is the global maximum. The origin of this fine structure is up to now unknown. It may arise from different active site structures, since slight changes in the structure around the SB can result in a shift of the absorption wavelength. It is possible that ChR2 has different stable ground state structures. Part of the present work is devoted to clarify the origin of this fine structure possibly allowing to propose a model of the binding pocket for ChR2.

Furthermore, the retinal configuration in the dark state is under debate as well. Resonance Raman spectroscopy studies measured at 293 K show C=C stretching bands at 1550 $cm^{-}1$ indicating an all-*trans* configuration while a band at 1557 $cm^{-}1$ indicates an additional presence of 13-*cis*, 15-*syn* retinal.[43] The mixture of both retinal configurations is also indicated by the fingerprint bands in the Raman spectra. Additionally, FT-IR difference spectroscopy at room temperature assigned negative bands, which show the ethylene vibration of all-*trans* and 13-*cis*, 15-*syn* retinal.[49] However, NMR solid state studies[50] found only a pure all-*trans* retinal in dark state ChR-2.

Central Gate, DC-Gate and Inner Gate

The crystal structure of ChR-2[31] forms a pore opened towards the extracellular side. It is confined next to the $RSBH^{+}$ by the *central gate* in the non-conductive state, which consists of a network of hydrogen bonds between the residues S63 (helix A), E90 (helix B), D253 (helix G) and N258 (helix G).[31,51] This region has also been found in the C1C2 chimera[52] (see Figure 2.1.2).

Moreover, an *inner gate* has been proposed separating the intracellular cavity in two cavities by the residues Y70 (helix A) E82, E83 (helix B), H134 (helix C), H265 (helix G) and R268 (helix G) (see Figure 2.1.2).[31]

Beside these two gates, which have been identified in the C1C2 chimera crystal structure as well,[30] the ChR-2 forms the extracellular gate separating the extracellular pore in two cavities. The residues M107, Q117, Y121, W124, S245 and H249 (helix C) separate the two cavities via an extended hydrogen bond network.

In this area of the protein an additional cation binding site has been found by theoretical studies[42] and an intrahelical hydrogen bond between residues E83 (helix B) and H134 (helix C) was detected.[51] The *central gate* and the *inner gate* contain the most functional residues, since mutation at these positions causes a change of the ion selectivity. In particular the mutations E90Q and H134R lead to a higher current in alkaline media and to a lower current in acidic conditions respectively.[53,54]

The E90 residue in ChR-2 has been found out by spectroscopic measurements to be protonated in the dark-state.[26,52,55] Furthermore, it has been suggested that E90 is involved

in the photocycle by the downward movement of the E90 side chain, which changes the hydrogen bonding network with N258 and S63 to E123 and D253. This was proposed to open the pore at the central gate in the early state of the photocycle.[34,47] After all, the E90 orientation in the crystal structure[31] is downward oriented in the initial dark state and builds a hydrogen bond network with E123, D253 and K93.

Another important part in the structure is the DC-gate, it consists of the residues C128 and D156. Mutations have shown, that exchange of those residues leads to a retarded photocycle in the stabilization of the open state and a delayed closing of the channel during the photocycle with respect to the WT.[39,56] Therefore, this part is important for the mechanism of the protein. These two residues are well conserved in all channelrhodopsins, which reveals their importance. However, double mutation stabilizes the photocycle in the gating state with little detectable return signal to the dark state.[57] Moreover, hydrogen bond interaction between D156 and C128 is relevant for the channel gating and the photocycle kinetics.[58] It connects helix C and D by a bridging water molecule.[31,42]

2.1.2 Photocycle of Channelrhodopsin-2

The all-*trans* to 13-*cis* photoisomerisation of the retinal chromophore is followed by a subsequent relaxation reaction back to the initial dark-state state within 20 s.[59,60] The further conformational changes of the protein happen through several intermediate states, which vary in conformation of the retinal, protein backbone conformation and protonation state of some residues. These rearrangements lead to the pore-opening (on-gating) and the followed permeation of cations (conducting states) through the membrane and a later closing of the channel (off-gating). Since the spectroscopic characteristic of the retinal is very sensitive to the surrounding protein, intermediates can be tracked and characterized by spectroscopic studies, like IR, UV/Vis and Raman spectroscopy. Additionally, conformational changes of the protein backbone or of amino acid side chains are measured by IR difference spectroscopy in the amide I region. However, the consequence of such changes, e.g. the conductance for mono- or divalent cations, have to be monitored by electrophysiological measurements.

Therefore, understanding the overall mechanism by which channelrhodopsins respond to light requires an identification and a subsequent characterization of all intermediates.

A photocycle model with four intermediate states P_1^{500}, P_2^{390}, P_3^{520} and P_4^{480} (Figure 2.1.3) was proposed based on time-resolved UV/vis, step scan FT-IR, tunable quantum cascade laser (QCL) IR spectroscopy studies[61] and electrophysical measurements.[43,62]

The intermediates P_1^{500} and P_2^{390} are supposed to consist of two sub-states which are not distinguishable by UV/vis spectroscopy measurements only by IR-spectroscopy measurements. The first intermediate (P_{1a}^{500}) is formed with a time constant of 3 ps.[35] It is

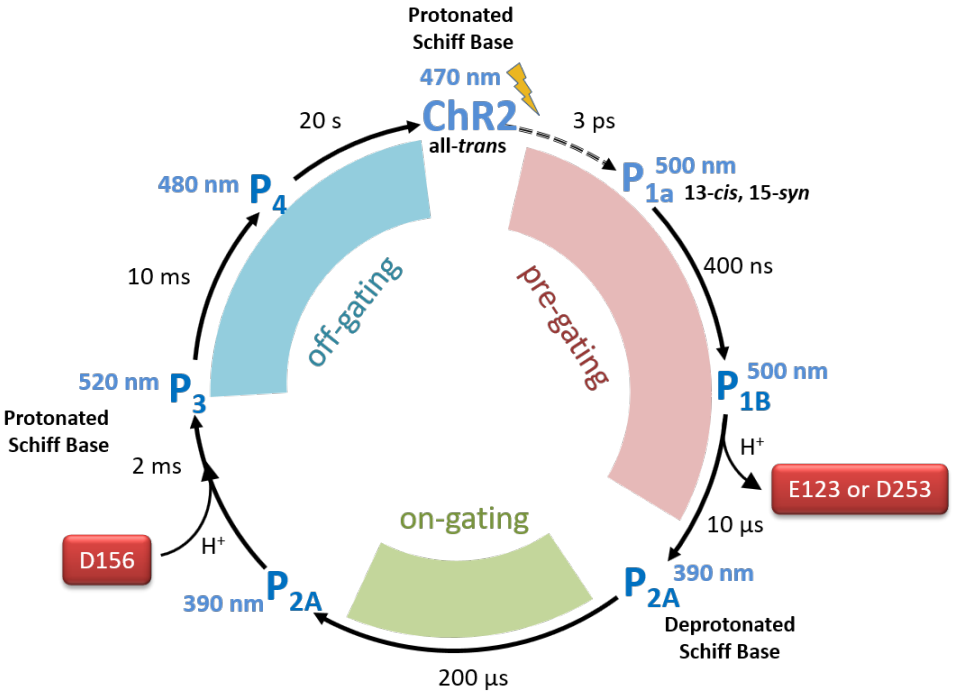


Figure 2.1.3: Proposed photocycle of ChR-2. The multi-exponential character of the intermediate formation is shown and the decay and the half-life time-constants $\tau_{1/2}$ are given. The super-scripted indices represent the absorption wavelength of the intermediate. The conductive states are the P_{2b} and P_3

assumed, that in this state the retinal isomerizes to the 13-*cis* conformation with subsequent adjustment of the binding pocket. The P_{1b}^{500} state arises with a time constant of 400 ns. Experimental results suggest, that with the arising of this state structural rearrangements, a change in the hydrogen bond pattern of the DC-Gate as well as conformational changes of the protein backbone occur.[61] The formation of the next state P_2^{390} ($\tau_{1/2}$ of 10 μs) can be explained by deprotonation of the SB.[35] However, the proton acceptor is still under debate. E123 has been proposed to act as primary proton acceptor[56,63] for the $RSBH^+$ proton, but time resolved step-scan FT-IR spectroscopy suggests D253 as proton acceptor.[39] However, studies from Kuhne *et al.* found out with time-resolved FT-IR, that both E123 and D253 are protonated in the P_{2a}^{390} intermediate.[47] Between $P_2^{390}a$ and $P_2^{390}b$ several protein backbone alterations takes place and lead to the opening of the channel (on-gating). The decay of the P_3^{520} intermediate is in concurrent with channel closing,[32,35] while the conducting state is present in P_{2b}^{390} and P_3^{520}.[59,63] The reprotonation of the SB is proposed to happen during the formation of P_3^{520} with a half-life of 2 ms, with D156 acting as proton acceptor.[39] In ChR2 the proton release to the bulk medium is coupled to the reprotonation of the SB.[43] The complete recovery of the channel occurs with $\tau_{1/2}$=20 s. The proton uptake takes place probably in the formation of P_3^{520} and P_4^{480} and corresponds to a two steps process, explaining the unusually long recovery time.[64] There is literature alternatively suggesting that the photocycle could be described by eight exponential terms, implying eight intermediate states.[39] However, at

this time only four intermediates havve been found by spectroscopic methods. There may be spectroscopically silent intermediate states or substrates which are spectroscopically similar to the identified intermediates.

The dark-state consists probably of all-*trans* and 13-*cis*, 15*syn* retinal, which lead to an additional photocycle. This photocycle consists of two intermediates I_1 and I_2 and has very fast kinetics with a complete recovery within 1 ms. However, whether this photocycle is functionally relevant, is not clear.[61]

Several mutational studies have been performed to investigate the structure function relationship and especially to expand the biological application of the ChR-2. Therefore, it has been found out, that the D253E mutant has a threefold slower decay of the P_2^{390} intermediate retreads channel closure, while the D253N mutant shows no channel behavior.[39] ChR-2-E123T accelerated (ChETA) mutant including E123T, E123A, E123Q mutants show smaller photocurrent amplitudes compared to the WT. They speed up the channel kinetics and are used for the stimulating of fast neuronal actions.[45,65]

2.2 Histidin Kinase Rhodopsin 1 (HKR1) - a Bimodal Switchable Rhodopsin

In *Chlamydomonas reinhardtii* it has been found, that besides the two Channelrhodopsins (*Cop3* and *Cop4*) and the two animal-related rhodopsins (*Cop1* and *Cop2*), there are at least four other Rhodopsins gene sequences (*Cop5* - *Cop8*).[66] The latter rhodopsins are linked to a histidin kinase (HK) and a response regulator (RR) receiver domain and opened up an unknown rhodopsin subfamily: the histidin kinase rhodopsins (HKRs). HKs are part of the two-component signaling system and operate as phosphorelay systems. They are usually coupled to sensor domains, whereby in HKRs the sensor is a light sensitive rhodopsin. In general the attached sensor of the HK detects a stimulus and induces the autophosphorelation of the HK. Hereafter the phosphate group is transferred to the RR receiver domain. This triggers conformational changes in the receiver domain, followed by activation of the RR effector domain. Thereafter, the RR effector domain initiate the cellular response to the signal. This procedure is also expected in HKRs, while only two of the four HKRs (*Cop5* and *Cop6*) additionally contain a C-terminal guanylyl-cyclase domain, which is supposed to act as the RR effector domain. Therefore, the signal transduction cascade and the cellular response is supposed to be initiated by the production cyclic GMP.[67]

In former studies the Cr-HKR1 (*Cop5*) rhodopsin domain was expressed, purified and spectroscopically investigated by the Hegeman group.[68] The characterization shows a surprising feature of two stable isoforms in the photocycle with different spectroscopic properties. Furthermore, with blue-green light or UVA-light it is possible to switch between

the two states. They are defined as Rh-BI (blue absorbing state) with an absorption maximum around 487 nm and Rh-UV (UV absorbing state) with an absorption maximum around 380 nm. The unusual stability of the two states over more than 24 h, especially the fact, that Rh-UV state contains a deprotonated retinal, has not been observed before and is highly interesting for development of new optogenetic tools.

As the different absorption wavelengths of the two stable states suggest, the states consist of different isoforms of retinal with a protonated (Rh-BI) or deprotonated (Rh-UV) Schiff base. The transition between Rh-BI and Rh-UV and vice versa is initiated by illuminating the protein with light of the absorption wavelength of the respective state. Thus, the retinal isomerizes and the protein undergoes several intermediate states by relaxing and rearranging the protein environment. As a final step in the photocycle the Schiff base protonation state changes and the protein ends up in the new state Rh-BI or Rh-UV, respectively. The photocycle of HKR1 was further characterized by pump-probe and laser-flash photolysis experiments[67,69] and is summarized in Figure 2.2.1.

Rh-BI to Rh-UV conversion

The excited state of Rh-BI state isomerizes and relaxes with a biexponential decay with time constants of 0.6 ps and 5 ps in the intermediate I1. This intermediate is red-shifted compared to Rh-BI and has an absorption maximum around 555 nm. Within 453 ps, I1 relaxes in the slightly blue-shifted state P550 with an absorption maximum around 550 nm. P550 has a life time of around 2-3 ms and leads to the P570 intermediate. In the last step the proton of the Schiff base is transferred to the counterion D239, therefore the P570 state has a much longer life time. After 27 ms the final state, Rh-UV, is reached and the first part of the photocycle is completed. This kind of reaction series is common in microbial rhodopsin photocycles.

Rh-UV to Rh-BI conversion

In contrast to the RH-BI to Rh-UV, the back reaction from Rh-UV to Rh-BI is quite particular. It starts with significant large excited state life time of 5 ps and 60 ps, within this time the isomerization happens and the only intermediate Rh-UV' shows up with related spectroscopic properties to the Rh-UV state. The transition from Rh-UV' to the final Rh-BI takes 2.1 ms with rearrangements in the protein and the reprotonation of the Schiff base.

The recent study by Luck[67] suggests two parallel photocycles and an additional isomerization around the C_{15}=N bond. Thus, Rh-BI and Rh-UV states consist of a mixture of 13-*trans*/13-*cis* and 15-*syn*/*anti* isomers.

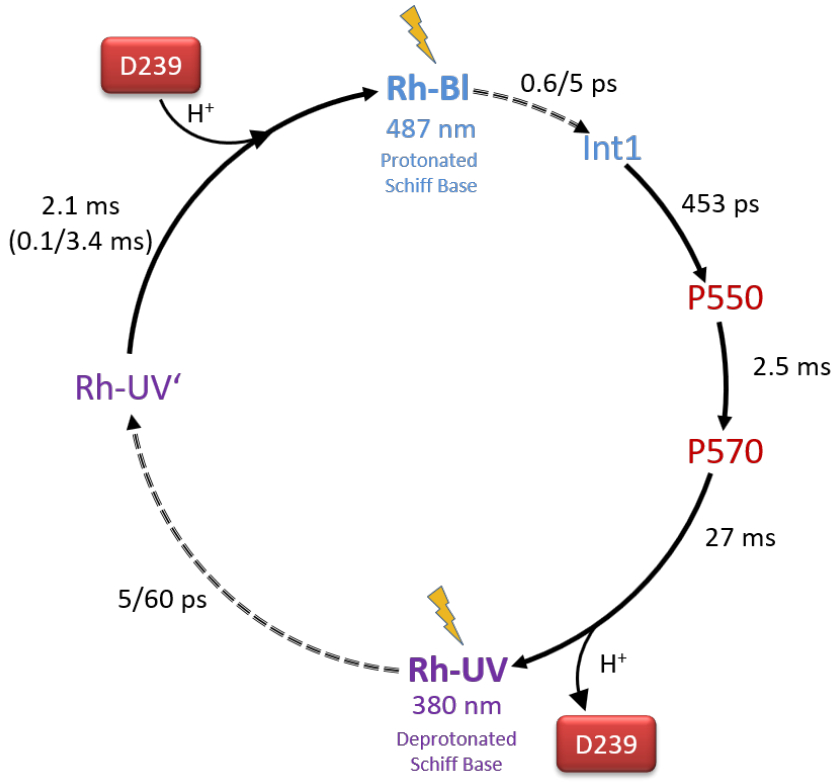


Figure 2.2.1: Photocycle of HKR1. Details see text.

2.3 Next Generation Optogenetic Tools

With the development of channelrhodopsins the field of neuroscience has been revolutionized in the last decade. The light-gated channel function of the ChR-2 and the high expression levels in neurons enables the genetic integration of channelrhodopsin in neurons. This offers a specific activation of the modified neurons in a large ensemble of neuronal cells only by light. This new technique is called Optogenetics and allows a fast targeting and precise control of events in biological systems. A functional analysis of cellular processes in health and disease can be possible by this technique. Optogenetics controls with short light flashes in a millisecond precision in time and local precision of 1 µm the specific activation of neuronal cells in neuronal tissues and can be applied from single cells to complex organisms like mammals.

The field of Optogenetics has been further developed and channelrhodopsin has been modified in order to increase the characteristics and fulfill the demands of the neuroscientific field. However, the inactivation of cells using moderate or low light intensities is not possible yet and has recently gained a lot of importance.

Part II

Methods

3 | Quantum Chemical Methods

The description of structures, properties or the thermodynamic nature of biological molecules by computational methods closes the gap between the experimental resolution and the atomistic scale. Their characterization and understanding, starting from the absorption of light in the femto second range up to conformational changes in the tertiary structure of a protein in the second range, span a wide scope of space and time and require different levels of accuracy (Figure 3.0.1).

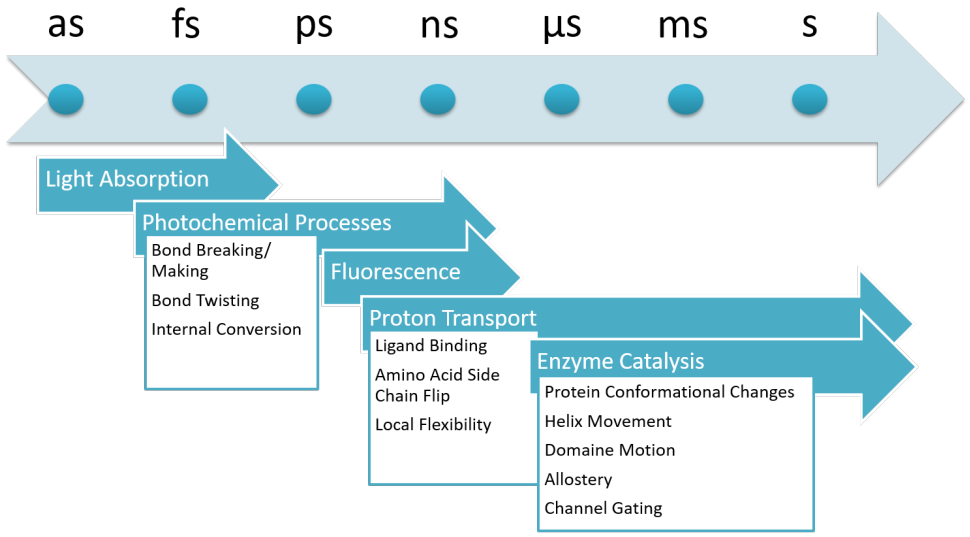


Figure 3.0.1: Timescale of optical processes and dynamic timescales of biological processes.

For example optical processes in biomolecules happen in ultrafast timescales in the range of as to fs and require an accurate and explicit description of the electronic structure during the process by quantum mechanical methods. In contrast, enzyme catalysis occurring in the range of several milliseconds up to seconds demands a less accurate description, but a rather extended simulation time (see Figure 3.0.1). Furthermore, for a correct description the inclusion of more than the protein, which leads to system sizes in order of 100 000 atoms is necessary. Here, the force field based or coarse grained methods solving Newton's equations of motion are well suited. However, for these complex systems MD simulations as

long as possible have to be performed, in order to ensure the characterization of the system in the equilibrium state.

Combining quantum mechanics with molecular mechanics (QM/MM) widens the possibilities for computational research and enables a more detailed description for example of optical processes within the protein environment.

The diversity of the eligible methods is huge and provides a large toolbox for different specifications and performances, on which several text books[70–73] have been written. However, this chapter will only give a brief overview about the methods used in this work.

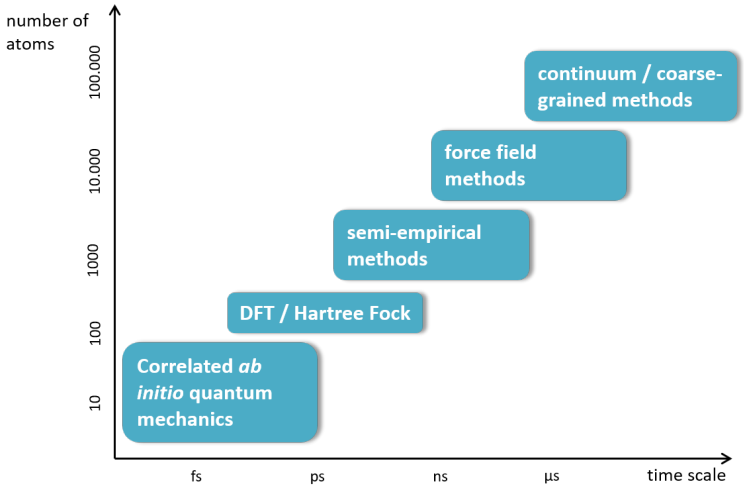


Figure 3.0.2: Relation between level of accuracy reached by the computational method and the describable system size

3.1 Schrödinger Equation and Hamiltonian

Due to the failure of classical physics in explaining phenomena like black-body radiation or Compton scattering, quantum mechanics was established in the last century. Its goal was to describe electrons, electronic properties and their behavior in matter in an accurate way. For this, the wave particle duality and the quantization of light had to be considered. At least, the combination of wave mechanics and the behavior of a particle having a momentum and a mass, brings up the time-dependent Schrödinger equation.

$$i\hbar\frac{\partial\Psi(r,R,t)}{\partial t} = \frac{\hbar^2}{2m}\nabla^2\Psi(r,R,t) + V(r,R,t)\Psi(r,R,t) \tag{1.1}$$

It is a differential equation with $\Psi(r,R,t)$ as the wave function dependent on the coordinates r for the electrons and R for the nuclei and it determines all possible solutions. On the left-hand side the operator $\frac{\partial\Psi(r,t)}{\partial t}$ describes the temporal development of the wave function and therefore it is called the time-evolution operator. On the right-hand side of Equation 1.1 the kinetic ($\frac{\hbar^2}{2m}\nabla^2$) and the potential energy ($V(r,t)$) build the Hamiltonian (Equation 1.2).

The Hamiltonian describes the total energy of the system and is the most important operator in quantum mechanics.

$$H = T + V = \frac{\hbar^2}{2m}\nabla^2 \ + V(r, R, t) \tag{1.2}$$

Considering $|\Psi(r, R, t)|^2$ in Equation 1.1, Max Born gave an interpretation of the wave function as the probability to find the system at a time t in the volume element dr.

For a free particle ($V(r, t) = 0$) the solution of the Schrödinger equation accords to the following wave function:

$$\Psi(r, t) = A \cdot e^{ikx - i\omega x} \tag{1.3}$$

In other systems the potential energy $V(r, R, t)$ determines the boundary conditions of the position-space wave function and is the only time-dependent variable on the right-hand side of Equation 1.1. Therefore, for a time-independent potential, the complex differential Schrödinger equation can be simplified. The wave function can be written as a product and the Schrödinger equation is simplified to the static or time-independent Schrödinger equation (Equation 1.4), which is an eigenvalue equation. There, the corresponding wave function Ψ gives the eigenstates and E gives the associated eigenvalues, which corresponds to the energy of the Hamiltonian.

$$E\Psi(r, R) = \left[\frac{\hbar^2}{2m}\nabla^2 + V(r, R)\right] \Psi(r, R) = H\Psi(r, R) \tag{1.4}$$

For a system with n electrons and N nuclei (Equation 1.5) the Hamiltonian consists of different operators, relating to the kinetic energy of the electrons ($\hat{T}_e$) and the nuclei ($\hat{T}_N$), the attractive interaction between the electrons and the nuclei ($\hat{V}_{eN}$) and the repulsion potential between electrons ($\hat{V}_{ee}$) and between nuclei ($\hat{V}_{NN}$).

$$\hat{H} = \hat{T}_e + \hat{T}_N + \hat{V}_{eN} + \hat{V}_{ee} + \hat{V}_{NN} \tag{1.5}$$

The Schrödinger equation can only be solved analytically for the H_2^+ molecule. Nonetheless, some approximations can be applied to solve systems with more than one nucleus and n electrons. Firstly, because the mass of the nuclei is orders of magnitude larger than the mass of the electrons, the motion of the nuclei can be separated from the motion of the electrons, so the nuclei can be treated as static. The wave function can be described as a product of the wave function of the electrons times the wave function of the nuclei.

$$\Psi(r, R) = \Phi(r; R)\chi(R) \tag{1.6}$$

This is called the Born-Oppenheimer approximation and the approximated Hamiltonian assumes the form:

$$\hat{H} = \underbrace{-\sum_{i}^{n} \frac{\hbar^2}{2m_e} \nabla_i^2}_{\hat{T}_e} \underbrace{-\sum_{i}^{n}\sum_{k}^{N} \frac{e^2 Z_k}{r_{ik}}}_{\hat{V}_{eN}} + \underbrace{\sum_{i<j}^{n} \frac{e^2}{r_{ij}}}_{\hat{V}_{ee}} + \underbrace{\sum_{k<l}^{N} \frac{e^2 Z_k Z_l}{r_{kl}}}_{\hat{V}_{NN}} \tag{1.7}$$

The nucleus-nucleus repulsion is then a constant effective potential and the wave function depends only on the coordinates of the electrons with the nuclei as parameters. This allows us to solve e.g. H_2^+ or He^+ molecules ions analytically. This so-called electronic Hamiltonian leads to an electronic Schrödinger equation (Equation 1.8), which is used as basis for all ground state quantum chemical methods.

$$H_{el}\Phi_{el} = E_{el}\Phi_{el} \tag{1.8}$$

Now, the aim of quantum chemistry is to find the best matching energy and wave function for the particular system. Since it is not possible to solve the Schrödinger equation analytically for systems containing more than three bodies, the variational principle is applied (Equation 1.9). It can be proved that any trial wave function Ψ results in a higher expected energy E than the exact energy E_0 of the system. Knowing that, the improvement of our trial wave function can be achieved by varying it until a energy minimum is reached.

$$E = \frac{\left\langle \Psi | \widehat{H} | \Psi \right\rangle}{\langle \Psi | \Psi \rangle} \geq E_0 \tag{1.9}$$

3.2 Hartree-Fock

The solution of the electronic Schrödinger equation for an n-electron system is still complex, because the wave function depends on the coordinates of n electrons. Applying the variational principle, the first step is to get an initial guess for the electronic wave function. Thus, the n-electron wave function Ψ can be constructed as the product of n one-electron wave functions Φ, the so called Hartree product (eq. 2.1). Here, every electron is described as a non-interactive particle.

$$\Psi(r_1, \ldots, r_n) = \frac{1}{\sqrt{M}} \phi_1(r_1) \cdot \phi_2(r_2) \ldots \phi_M(r_n) \tag{2.1}$$

The one-electron functions Φ are the so called molecular orbitals (MO) and can be built via linear combination of atomic orbitals (LCAO):

$$\phi_i = \sum_j c_j^i \eta_j \tag{2.2}$$

The Hartee product considers only the spatial degrees of freedom of the electrons, but the additional degree of freedom, the spin coordinate is neglected. The spin coordinate can take two values, either spin up α or spin down β. Therefore, every orbital can be occupied

by 2 electrons and the amount of orbitals is reduced to $M = \frac{n}{2}$. Unfortunately, this simple approach violates the Pauli exclusion principle and enables electrons to be consistent in all quantum numbers. To overcome this issue, the wave function has to be antisymmetric. This can be achieved by constructing the n-electron wave function as a Slater-determinant:

$$\Psi(r_1, \ldots, r_n) = \frac{1}{\sqrt{M!}} \begin{vmatrix} \phi_1(x_1) & \phi_2(x_1) & \cdots & \phi_M(x_1) \\ \phi_1(x_2) & \phi_2(x_2) & \cdots & \phi_M(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_1(x_n) & \phi_2(x_n) & \cdots & \phi_M(x_n) \end{vmatrix} \tag{2.3}$$

With this initial guess all c_j^i in Equation 2.2 will be varied in order to improve the MOs $\phi_M(x_n)$ in Equation 2.3 until the minimum is found. Thereby the energy of the system will never be below the energy of the ground state E_0 according to the variational principle (Equation 1.9). The orthonormality of the MOs ϕ_i has to be conserved for every step of the variation:

$$\langle \phi_i | \phi_j \rangle = \delta_{ij} \begin{cases} \delta_{ij} = 1 \; if \; i = j \\ \delta_{ij} = 0 \; if \; i \neq j \end{cases} \tag{2.4}$$

Minimization of the wave function requires the solvation of the eigenfunction in an iterative way, in the so-called self-consistent field, due to the fact that all electrons interact with each other while the Hamiltonian itself is dependent on this interaction, the wave function can not be determined until the Hamiltonian is known.

The Hamiltonian in Equation 1.7 can be divided to the one-electron operator h_i, which describes the motion of one electron i in the field of all nuclei and the two-electron operator g_{ij}. This describes the repulsive interaction between two electrons (Equation 2.5).

$$\begin{aligned} h_i &= -\frac{\hbar^2}{2m_e}\nabla_i^2 - \sum_i^n \sum_k^N \frac{e^2 Z_k}{r_{ik}} = T_e - V_{Ne} \\ g_{ij} &= \frac{e^2}{r_{ij}} = V_{ee} \\ H_e &= \sum_i^n h_i + \sum_{i<j}^n g_{ij} + V_{NN} \end{aligned} \tag{2.5}$$

If H_e for one electron is now inserted into the energy equation $E = \left\langle \Psi | \widehat{H} | \Psi \right\rangle$, the two terms for the repulsive potential in Equation 2.6 will arise. The J_{ij} is called the Coulomb Integral and the K_{ij} is called the exchange integral, which results from the antisymmetry operator in the Slater-Determinant. The Coulomb integral represents the classical repulsion between two electron densities, while the exchange integral is a complete quantum mechanical effect and has no classical analogon. It can be interpreted by the statement that two electrons with the same spin never stay at the same position.

$$
\begin{aligned}
E = & \sum_{i}^{n} \langle \phi_i(1) | h_1 | \phi_i(1) \rangle + \\
& \sum_{i}^{n} \sum_{j>i}^{n} (\langle \phi_i(1)\phi_j(2) | g_{12} | \phi_i(1)\phi_j(2) \rangle - \langle \phi_i(1)\phi_j(2) | g_{12} | \phi_j(1)\phi_i(2) \rangle) + V_{nn} \\
E = & \sum_{i}^{n} h_i + \sum_{i}^{n} \sum_{j>i}^{n} (J_{ij} - K_{ij}) + V_{nn}
\end{aligned}
\tag{2.6}
$$

Now, in order to minimize the energy of the MOs ϕ_i Equation 2.6 has to be varied under orthonormal constrains. For this constrained minimization, the *Langrangian mulitpliers* are the method of choice and results in the Hartree Fock equation:

$$\hat{f}\phi_i(x_i) = \epsilon_i \phi_i(x_i) \tag{2.7}$$

with the Fock operator expressed as:

$$f_i = \underbrace{-\frac{1}{2}\nabla_i^2 - \sum_{i}^{N} \frac{Z_i}{r_{ij}}}_{\hat{h_i}} + \sum_{ij}^{n} \underbrace{(\hat{J}_{ij} - \hat{K}_{ij})}_{\hat{v}_{HF}} \tag{2.8}$$

The Fock operator is an effective one-electron energy operator. It describes in the first term on the right hand side of Equation 2.8 ($\hat{h}$) the motion of one electron in the potential field of of all nuclei and the repulsion of all other electrons via J_{ij} and K_{ij} operators.

By the Hatree-Fock approach, the complex interaction between n-electrons is approximated by the Hartree-Fock potential $\hat{v}_{HF}$, corresponding to the averaged potential of n-1 electrons, which is felt by a single electron.

To determine the averaged Hartree-Fock potential the MOs have to be known, so the equation can not be solved linearly. Rather it needs to be solved iteratively in a self-consistent field (SCF), starting from an initial guess which is iteratively enhanced at every step. The Hartree-Fock method does not require the use of empirical parameters, making it an *ab initio* method.

3.3 Electron Correlation and Excited States

In Hartree-Fock methods not all important interactions are considered. Firstly, there is only an averaged interaction between the electrons, that means one electron feels the other electrons as an averaged electron cloud. Additionally, only the occupied orbitals are considered in the Slater-Determinant, determining the wave function. Thus, we neglect one important contribution to the chemical bond, the correlation interaction. It corresponds to the motion of the electrons being correlated in the orbitals. Indeed, the electrons have more

space to move, due to the neglected unoccupied orbitals. Rather, the electrons are further apart from each other on average and the repulsive interaction between them is reduced, which contributes to the total energy of the molecule. The correlation interaction contributes only with $\sim$1 % to the total energy, however, it is important in order to describe chemical phenomena.

A possible strategy is to consider the unoccupied orbtitals in the wave function by adding Slater-Determinants, containing the virtual molecular orbitals, to the Hartree-Fock Slater-Determinant (Equation 3.1). They are superpositions of ground state and excited state determinants and are also called *Configuration State Function* (CSF). The occupied orbitals are replaced by virtual or unoccupied orbitals of the reference Hartree-Fock Slater-Determinant Φ_{HF}.

$$\Psi = a_0\Phi_{HF} + \sum_i a_i\Phi_i \tag{3.1}$$

3.3.1 Configuration Interaction

One approach is to add the Slater-Determinants by a linear combination, this is called the *Configuration Interaction (CI)* method and is the easiest way to include the correlation of the electrons in the calculation. Each Slater-determinant in Equation 3.2 represents one electron configuration, where a occupied molecular orbital (MO) is replaced by a virtual, unoccupied MO, thus the virtual Orbitals are included in the calculation. The singly, doubly and triply "excited" determinants relative to the HF determinant refer to the Single (S), Double (D) and Triple (T) Determinants etc. up to N_{elec} excited electrons. Since the MOs are build out of atomic orbitals (basis functions) the number of virtual MOs is dependent on the number of basis functions and the electrons. This means, the larger the basis set, the more virtual MO can be constructed and considered in the many-electron wave function and the more correlation energy of the electrons is considered. Therefore, the accuracy of the CI method can be improved by increasing the size of the basis set. The consideration of all possible excited Slater determinants and an infinitely large basis set will give the exact excitation energy within the Born-Oppenheimer approximation and is called *Full CI*. Unfortunately, this will lead really fast to very expensive calculations and is only feasible for small molecules. For larger systems the considered electron excitations are limited and only single excitations (CIS), double excitations (CID) or a combination of them (CISD) are used.

$$\Psi^{CI} = a_0\Phi_{HF} + \sum_S a_S\Phi_S + \sum_D a_D\Phi_D + \sum_T a_T\Phi_T + ... \tag{3.2}$$

The expansion coefficients (a_S, a_D a_T, etc.) of the determinants can be obtained by minimizing the energy of the system. The usual way in quantum chemistry is the application

of the variational principle to minimize the energy. The expansion coefficients of the Slater-Determinants (a_S, a_D, a_T) will be varied in order to minimize the energy, while the CSF (Φ_S, Φ_D , ...), representing the excited wave-function obtained by Hartree-Fock calculations ($\Phi_S = \sum_j c_j^i \eta_j$) expanded in a set of basis functions, will held fixed and the expansion coefficients of the MOs (η_j) wont be optimized. This evaluation leads to the CI matrix, where the energies of the respective states are given by the diagonal elements of the matrix, while the expansion coefficients represent the contribution of the Slater-Determinants to the energy of the respective state.

3.3.2 Multi-Configuration Self-Consistent Field (MCSCF) and Multi-Reference Configuration Interaction (MRCI)

As mentioned above, the MO coefficients (η_j) of the CSF are fixed in the CI calculation. But the method can be further improved by additionally optimizing this coefficients within the calculation. This leads to the *Multi-Configuration Self-Consistent Field* method (MCSCF). Since the amount of coefficients which needs to be optimized increases rapidly with the amount of electrons and orbitals, requires a limitation of the optimized MOs. Only the electronic configurations of the orbitals, which have the highest contribution to the property of the interesting state, should be considered in the calculation. Therefore, the MOs, which should be optimized in the calculation, are defined by a user-selected active space, while the coefficients of the MOs in the inactive space wont be optimized and remain on the HF-level. In the active space, MOs with a varying occupation number are included in the active space and MOs, which are fully occupied or fully empty are considered as inactive. The size of the active space and the selection of the right MOs are crucial in order to include all chemically relevant orbitals. This requires insights into the photochemical and photophysical characterization of the molecule.

When considering all electronic configurations within the orbitals of the defined active space it is the so-called *Complete Active Space Self-Consistent Field* (CASSCF). This is the most attractive method (see Figure 3.3.1).[74] Whereas the *Restricted Active Space Self-Consistent Field* (RASSCF) is again an approximation of the CASSCF, since the considered electronic configurations are restricted excitations from RAS1 to RAS3, while in the RAS2 all configurations are allowed. Usually all π-Orbitals of the molecule need to be included in the active space. Therefore the RASSCF method can be very useful for large chromophor systems like Chlorophyll containing more than 6 double bonds, since the CASSCF is limited to an active space of 12 electrons in 12 orbitals, due to the computing power.

Up to here, the Slater-Determinants are generated based on the single Hartree-Fock wave function. However, further improvement of the method is gained by describing the systems

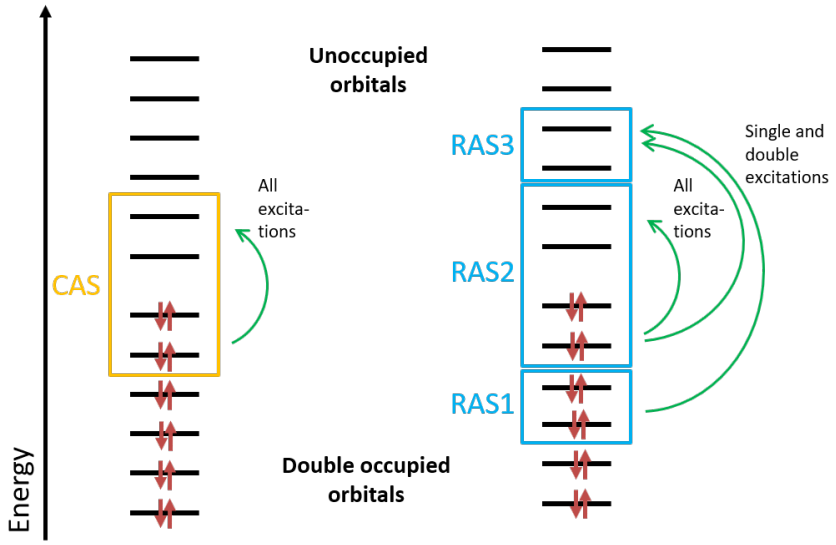


Figure 3.3.1: Illustration of CAS and RAS orbital partitions. In the CAS (left) all possible electronic configurations are allowed within the determined orbitals. In RAS (right) only limited excitations are allowed from RAS1 to RAS3, while in RAS2 the configurations are unlimited within the selected orbitals.

with a MCSCF wave function as a reference. Thus, the *Multi-Reference CI* (MRCI) approach considers more than one reference wave function, including different excitation levels. This offers a balanced description of the states.

The CI methods are not consistent in size, i.e. the calculated energy of two separated twin systems is not the same as the energy of the twin systems described in one calculation. For example, twin systems each with two electrons are in the separated calculation double excited monomers, while the combined twin system formally correspond to quadruple excitations in the dimer. This inconsistency in size is approximately corrected by the Davidson correction.

$$E_{CISDTQ} \approx E_{CISD} + \Delta E_Q \Delta E_Q = (1 - C_0^2) E_{CISD} - E^{HF} \tag{3.3}$$

3.3.3 Coupled Cluster

In contrast to CI methods, an exponential ansatz is chosen to generate the wave function in the Coupled Cluster (CC) methods. Here, the *Full CI* wave function is described by:

$$\Psi = e^T \Psi \tag{3.4}$$

where T is the cluster operator, which generates the possible excited Slater-determinants and is given by:

$$T = T_1 + T_2 + T_3 + ... + T_n \tag{3.5}$$

If $1 + T$ is applied to the Hartree-Fock wave function, one will get the *Full CI* wave function. The difference in Coupled Cluster to CI methods is the exponential ansatz and this ansatz has an enormous advantage when truncating T. Higher excitations are considered in the calculations, due to the Taylor expansion of the exponential function. For example the wave function for CCD method, where only double excitations are considered is given by:

$$\Psi_{CCd} = (1 + T_2 + \frac{T_2^2}{2!} + \frac{T_2^3}{3!} + ...)\Psi_{HF} \tag{3.6}$$

The $\frac{T_2^2}{2!}$-term contains quadruple excitations, since it is the product of two double excitations. While the $\frac{T_2^3}{3!}$-term contains up to hextuple excitation due to a multiplication of three double excitations. This includes exactly the missing excitation in the CI method and makes the CC method more exact and size consistent.

3.4 Density Functional Theory

The proof by Hohenberg and Kohn[75] established, that the ground state energy can be completely defined in an alternative way by the electron density ρ. With this theorem the complex wave function, dependent on 4N coordinates (three spatial and one spin coordinate for each of the N electrons) in the wave function based methods described above, can be reduced to a functional of the electron density, which is dependent on only three coordinates. Therefore, the energy $E[\rho(r)]$ and the properties of the system is uniquely defined by the electron density ρ. However, the functional, which connects the energy with the electron density is not known and needs to be designed. This is the goal of the DFT method.

The basis of this method is explained by the two Hohenberg-Kohn theorems. In the first theorem, the electron density is defined by the external potential of the nuclei potential and corresponds to electron density of the ground state for the lowest energy of the system, which is determined in the second Hohenberg Kohn theorem. To identify the ground state, the electron density needs to be varied untill the energy reaches a minimum.

The energy of the system (Equation 4.1) can be divided in the kinetic energy of the electrons ($T[\rho]$), the electron-nucleus interaction ($V_{eN}[\rho]$) which can be described with a classical Coulomb interaction, the classical nuclear-nuclear repulsion $V_{NN}[\rho]$ and the interaction between two electrons $V_{ee}[\rho]$. The electron-electron interaction is similar to the Hartree-Fock method (compare section 3.2), composed of the classical Coulomb interaction $J[\rho]$ and a non classical exchange-correlation functional $E_{xc}[\rho]$ (see Equation 4.2).

$$E[\rho] = T[\rho] + V_{eN}[\rho] + V_{ee}[\rho] + V_{NN}[\rho] \tag{4.1}$$

$$V_{ee}[\rho] = J[\rho] + E_{xc}[\rho] \tag{4.2}$$

The exchange-correlation functional $E_{xc}[\rho]$ represents the exchange interaction for electrons with the same spin (exchange correlation) and the correlation interaction of electrons with opposite spin (coulomb correlation) as well as the self-interaction of the electrons. Due to the Born-Oppenheimer approximation the nuclei are considered as infinitely heavy compared to the electrons and the velocity of the nuclei can be neglected. Therefore, the nuclear-nuclear repulsion $V_{NN}[\rho]$ accounts to the total energy only as constant value. The attraction between electrons and nuclei ($V_{eN}[\rho]$) and the Coulomb Interaction ($J[\rho]$) can be calculated by their classical expressions as this was shown for the Hartree-Fock method. However, for the exchange-correlation functional $E_{xc}[\rho]$ and for the kinetic energy $T[\rho]$ the shape of the functional is not known, but can be approximated by several approaches.

Early approaches known as *Thomas-Fermi* and *Thomas-Fermi-Dirac* model tried to depict the unknown terms as functional of the electronic density. However, the description of the electron density as an uniform electron gas has been found to give poor energies for molecules, but performs well for the valence electrons in certain periodic metalic systems (like in solid-state physics).

The idea of Kohn and Sham[76] to solve this complex problem was to approximate the electron density ρ in a auxiliary set of n non-interacting wavefunctions (Equation 4.3), the Kohn-Sham orbitals Φ_i. In other words, the idea from the wave function methods, using molecular orbitals, is re-introduced for the DFT method to describe the electron density (Equation 4.3).

$$\rho(r) = \sum_{i}^{n} |\Phi_i|^2 \tag{4.3}$$

Here, it is assumed that a ground state electron density of non-interacting electrons is the same as the electron density of a system with interacting electrons. Indeed, in reality the electrons interact with each other and the neglected energy term, due to the non-interacting approximation, is absorbed to the exchange-correlation term $E_{xc}[\rho]$. So the Schrödinger equation can be expressed as a set of equations with the Kohn-Sham orbitals Φ_i in an effective potential. These equations are called Kohn-Sham equations (eq. Equation 4.4) and have to be solved to obtain the ground state energy.

$$\left[-\frac{\hbar^2}{2m}\nabla^2 + v_{eff}(\rho)\right]\Phi_i = \epsilon_i \Phi_i \tag{4.4}$$

The first term on the left hand side is the kinetic energy of the electrons. The Kohn-Sham equations describe a set of moving non-interacting electrons in a potential field given by the effective potential $v_{eff}(\rho)$ (eq. Equation 4.5).

$$v_{eff}(\rho) = \underbrace{-\sum_{i}^{n}\sum_{k}^{N}\frac{eZ_k}{r_{ik}}}_{V_{eN}} + \underbrace{\frac{e^2}{4\pi\epsilon}\int\frac{\rho(r_j)}{r_{ij}}d^3r_j}_{V_{ee}} + \underbrace{\sum_{k<l}^{N}\frac{Z_k Z_l}{r_{kl}}}_{V_{NN}} + \underbrace{\frac{\partial E_{xc}}{\partial\rho}}_{V_{xc}} \tag{4.5}$$

The effective potential $v_{eff}(\rho)$ is dependent on the electron density itself, therefore the Kohn-Sham equations have to be solved iteratively. It includes the attraction potential between the nuclei and the electrons $V_{eN}[\rho]$, the classical electron-electron repulsion V_{ee}, the classical nuclear-nuclear repulsion V_{NN} and the non-classical exchange-correlation energy. The exchange-correlation potential accounts for the exchange and correlation effects of the electrons in the total energy. If the exact exchange potential is known, all correlation effects between the electrons are considered in the calculations and the energy could be described exactly within the Born-Oppenheimer approximation. But in fact, the exchange correlation is unknown and has to be approximated. For this approximation several approaches have been determined in the past. The sum of the electron energies can than be calculated as:

$$\sum_{i}^{n} \epsilon_i = \sum_{i}^{n} \left\langle \Phi_i \left| -\frac{\hbar^2}{2m}\nabla^2 + v_{eff}(\rho) \right| \Phi_i \right\rangle = T[\rho] + \int v_{xc}\rho \, dr \tag{4.6}$$

$$T[\rho] = \sum_{i}^{n} \epsilon_i - \int v_{xc}\rho \, dr \tag{4.7}$$

For the Kohn-Sham orbitals atomic orbitals from a basis set as described in the Hartree-Fock method are used, and a matrix formulation of the eigenvalue is obtained by applying the variational principle.

The total energy can be written in an alternative way as:

$$E[\rho] = \sum_{i}^{n} \epsilon_i - \int v_{xc}\rho \, dr - \frac{1}{2}\int\int \frac{\rho(r_j)\rho(r_i)}{|r_i - r_j|} d^3r_j d^3r_i + E_{xc} \tag{4.8}$$

3.4.1 Exchange Correlation Functionals

For calculations with DFT it is necessary to choose an adequate basis set for the orbitals and additional functionals to describe the exchange-correlation energy. Since the exchange-correlation energy is the difference to the exact energy, the quality of the DFT calculations depends on the complexity of the functionals.

A simple approach to consider the exchange-correlation energy is the local density approximation (LDA), where the functional depends only on the electron density. It is able to describe accurately vibrational frequencies and geometries, but it cannot be applied to obtain reaction energies.

In this respect, the generalized gradient approximation (GGA) is a better choice, since it additionally depends on the gradient of the density (example are PBE[77] or BLYP[78,79]). An additional improvement can be obtained by combining two types of functionals as this is the case of B3LYP[80] where the exchange part is calculated with the Hartree-Fock formalism, which is called the Hartree-Fock exchange. Therefore, the improved results due to these functionals arise from an error cancellation: DFT overestimates the delocalisation of the

orbitals but Hartree-Fock underestimates it. Mixing these two effects results in a remarkable improvement.

The DFT method fails when describing charge transfer processes, where formally an electron is moved over a larger distance in a transition process like for example the absorption. This is due to the fact, that the exchange correlation functionals are by definition local functionals depending on the density or their derivatives. Furthermore the exchange correlation is not described in the right way, since the functions usually used to describe it are local in nature and exchange correlation is a delocalized effect, which is described correctly by the Hartree-Fock method. In order to improve the functionals, for example for charge transfer transitions like they happen in retinal molecules, the range-separated functionals have been designed. Here, the electron-electron interaction is partitioned in a short-range interaction and in a long-range interaction, by a complementary error function. The short-range interaction can be described by the DFT method, which performs well for short-range interactions and the long-range interactions are described by the exact Hartree-Fock exchange. Drawback of this approach is the increased complexity of the calculations and the consequential reduced computing efficiency.

3.4.2 Linear Response Time-Dependent Density Functional Theory (TD-DFT)

The Linear Response Time-Dependent DFT is the most prominent and widely used approach for the calculation of excitation energies of medium sized to large systems. The method is based on the Runge-Gross theorem stated in 1984,[81] which is the analog to the Hohenberg-Kohn theorem. Based on this theorem the electron density $\rho(r,t)$ at an arbitrary time uniquely determines the time-dependent external potential $V_{ext}(r,t)$. This makes it possible to calculate the time-dependency of many-electron system by using the electron density.

The excited state properties can then be calculated by the linear response formulation. It approximates the electronic transition from the ground state to the excited state by a first order perturbation of the electronic density in the ground state, under the assumption that the perturbation is slowly tuned on (adiabatic approximation). The perturbation of the electronic density results in the energy of the excited state and the oscillator strength for this state, while an explicit description of the electron density in the excited state is not received. Since TD-DFT can be considered as a perturbation of the DFT method, the advantages and drawbacks of the DFT methods and their functionals are still present. Therefore, TD-DFT is not suitable for charge transfer transitions like in the retinal molecule.

3.5 Non-Adiabatic Dynamic Simulations

For the theoretical investigation of ultrafast non-adiabatic photo-induced processes an excited state dynamic on different electronically states is required. The potential surface of the ground and the excited states needs to be treated in a very accurate manner, while simultaneously the propagation of the system evolves along the surface. This is a challenging task for theoretical methods. The combination of semi-classical dynamics with computationally quantum methods allows the efficient calculation of electronic structure propagation in different electronic states. In other words, classical trajectories are simulated on a single potential surface (PES), at which the nuclei are described by newtons equation of motion and the electrons are treated by quantum chemical methods.

The crossing of two potential surfaces or the characterization of the radiationless transition from the excited state back to the ground state requires a description which is consistent with the QM methods. Tully's surface hopping approach[82,83] lends itself to combine the semi-classical dynamic simulation with the quantum chemical methods for electronic structure calculations.

Therefore, the properties of dynamic systems in the excited state can be described approximately by averaging over an ensemble of classical trajectories.

The Born-Oppenheimer approximation is basis for modern quantum chemistry and is the fundamental approximation for semi-classical dynamics, since it enables the separation of the electronic wave function and the nuclear wave function. So, the nuclei can be propagated separately from the electrons. However, at a crossing point of two adiabatic surfaces, the Conical Intersection, the wave functions of the two states couple with each other and the Born-Oppenheimer approximation breaks down at this point. Therefore, the surface crossings are only approximately described and the transition between the states happens via a hopping algorithm. This means, when two potential surfaces come close to each other the probability $P_{1\to 0}$ for a transition between the surfaces increases and can be calculated based on the Landau and Zener formular Equation 5.1.[84,85]

$$P_{1\to 2} = e^{-\pi \frac{\Delta E_{12}^2}{h\nu |\dot{\Psi}_1 - \dot{\Psi}_2|}} \tag{5.1}$$

where ΔE_{12} is the energy gap between the surfaces and ν is the velocity of the wave packet, while

$$\dot{\Psi}_i = \frac{dU_i}{dQ} \tag{5.2}$$

is the continuation of the diabatic state across the adiabatic crossing along the reaction coordinate Q.

If the probability for a hop is high, the system hops to the other potential surface and the propagation is continued. This could happen when the velocity of the system is high, the two

potential surfaces come very close to each other or if the shape of the adiabatic surfaces change abruptly in the region of the avoided crossing. Since the hopping probability needs to be evaluated on each time step, the value of the time step is crucial for the description of the surface hopping, especially when concerning relaxation times of the excited state. It is usually determined to less than 1 fs

The Tully Surface Hopping approach is the most prominent method for non-adiabatic dynamics and is implemented in almost all quantum chemical methods, which are able to describe excited state dynamics. It combines surface hopping with *on-the-fly* dynamics and is therefore very efficient.

4 | Semi-Empirical Methods

The quantum mechanical methods described above are complex and sophisticated, but computationally very costly. Calculations on this high level of methods are limited to small systems with a system size less than hundred atoms. Even on such small systems the computation time amounts several hours, if not days. Therefore, only single calculations on optimized structures are possible to perform. But processes in nature happen over extended timescales in complex environments and moreover dynamical aspects are necessary to understand the characteristics of these processes. For that reason, it is not enough to consider single molecules in single calculations to get the whole picture of the properties of a system. Unfortunately, the *ab initio* quantum mechanical methods are computationally too expensive to consider all these effects and the propagation with most of the methods is computationally too expensive for trajectories in a ns timescale. Further assumptions are needed on the known concepts in order to simplify the calculations and reducing the computation time, with a simultaneously conservation of the accuracy of the quantum mechanical methods, so that the treatment of the electrons and their correlation does not lead to artificial errors due to the method.

4.1 Density-Functional Tight Binding

The density-functional tight binding method (DFTB) is a semi-empirical method based on the Density Functional Theory (DFT). It originates from solid-state physics and is commonly used to describe crystals with a rigid structure. The approximation which is made reduces the basis set and includes only the valence electrons of a molecule, while treating the core electrons with two-center potentials. Furthermore, the Hamiltonian and the overlap matrix for these electrons are parametrized for a grid of distances. One of the time consuming steps, the evaluation of all off-diagonal elements of the overlap matrix, is thus eliminated.

In the tight-binding theory, the atoms are highly localized in space and the electron density can be approximated by a reference density ρ^0 and by the fluctuations around this reference

$\partial\rho$. The reference density is expressed as a sum of neutral atomic densities ρ_k^0.

$$\rho(r) = \rho^0(r) + \partial\rho(r) = \sum_k \rho_k^0 + \partial\rho \tag{1.1}$$

The DFTB energy can be obtained by insertion of the approximated electron density in the DFT total energy Equation 4.8 and can then be written as:

$$\begin{aligned}
&E[\rho^0(r) + \partial\rho(r)] = \\
&\sum_i \left\langle \Phi_i \left| -\frac{\hbar^2}{2m}\nabla^2 + V_{eN} + \int \frac{(\rho_j^0 + \partial\rho_j)(\rho_i^0 + \partial\rho_i)}{|r_i - r_j|} d^3r_j + V_{xc}[\rho^0 + \partial\rho] \right| \Phi_i \right\rangle \\
&- \frac{1}{2}\int\int \frac{(\rho_j^0 + \partial\rho_j)(\rho_i^0 + \partial\rho_i)}{|r_i - r_j|} d^3r_j d^3r_i \\
&- \int v_{xc}[\rho^0 + \partial\rho](\rho^0 + \partial\rho)\, dr + E^{xc}[\rho^0 + \partial\rho] + E^{NN}
\end{aligned} \tag{1.2}$$

The exchange-correlation energy $E_{xc}[\rho^0 + \partial\rho]$ is expanded in a Taylor series around ρ^0 and the terms depending on ρ^0 are separated. This leads to a new formulation of the energy:[86–88]

$$\begin{aligned}
E^{DFTB} =& \sum_i \left\langle \Phi_i \underbrace{\left| -\frac{\hbar^2}{2m}\nabla^2 + V_{eN} + \int \frac{\rho_j^0}{|r_i - r_j|} d^3r_j + V_{xc}[\rho^0] \right|}_{\hat{H}^0} \Phi_i \right\rangle \\
&- \frac{1}{2}\int\int \frac{\rho_j^0 \rho_i^0}{|r_i - r_j|} d^3r_j d^3r_i - \int v_{xc}[\rho^0]\rho^0\, dr + E^{xc}[\rho^0] + E^{NN} \\
&+ \frac{1}{2}\int\int \left(\frac{1}{|r_i - r_j|} + \left.\frac{\partial^2 E^{xc}[\rho]}{\partial\rho\partial\rho_j}\right|_{\rho_i^0, \rho_j^0} \right) \delta\rho_i \delta\rho_j\, dr_i dr_j \\
&+ \frac{1}{6}\int\int\int \left.\frac{\partial^3 E_{xc}[\rho]}{\partial\rho_i\partial\rho_j\partial\rho_k}\right|_{\rho_i, \rho_j, \rho_k} \delta\rho_i \delta\rho_j \delta\rho_k\, dr_i dr_j dr_k + \dots \\
&= E^0[\rho^0] + E^1[\rho^0, \partial\rho] + E^2[\rho^0, (\partial\rho)^2] + E^3[\rho^0, (\partial\rho)^3] + \dots
\end{aligned} \tag{1.3}$$

where $\hat{H}^0$ represents the energy (Hamiltonian) of the reference density ρ^0.

The truncation after different terms of the Taylor series constitutes different DFTB methods, while truncating after the second term ($E^1[\rho^0, \partial\rho]$) results in the original DFTB method.[89,90] DFTB2[87] considers also the second-order term ($E^2[\rho^0, (\partial\rho)^2]$) and DFTB3[88] includes additionally the third-order term ($E^3[\rho^0, (\partial\rho)^3]$).

The terms for the electron interaction and the exchange-correlation energy ($E^0[\rho^0]$) as well as the nucleus-nucleus repulsion are summarized by E_{rep}, a sum of pairwise potentials which are approximated by atom types and distances taken from several references. This leads to the neglection of the three-center contributions.

$$E_{rep} = E^0[\rho^0] = \frac{1}{2}\sum_{ab} V_{ab}^{rep}[\rho_a^0, \rho_b^0, r_{ab}] \tag{1.4}$$

Thus, the total energy of the standard DFTB method is given by (here the Taylor series is truncated after the first term):

$$E_{DFTB}[\rho(r)] = \sum_i \epsilon_i + E_{rep} \tag{1.5}$$

The density fluctuation $\partial\rho$ is neglected in this method and this represents the origin of the method from the solid state physics tight-binding method, which is usually applied on nonpolar, homologous systems or systems with no charge transfer, e.g. crystals. Here the second term has no effect on these systems. The drawback of standard DFTB method is the neglected second term, without this term the description of polar systems with heterogeneous electron densities and charge transfer is not reasonable, which are crucial effects in molecules.

DFTB2 considers the second-order term, where the density fluctuation is described as an atomic contribution and the total charge of the system is given by the sum of the charge fluctuations (δq) of all atoms:

$$\int \partial\rho = \sum_k \delta q_k \tag{1.6}$$

$$\delta q_k = q_k - q_k^0 \tag{1.7}$$

q_k is the Mulliken charge and q_k^0 the number of valence electrons of the atom. The second-order term can be written as:

$$E_{2nd} = \frac{1}{2}\sum_{ab} \delta q_a \delta q_b \gamma_{ab} \tag{1.8}$$

where γ_{ab} represents the second-order approximation, i.e. the fluctuation of charges depend on distances, therefore the Coulomb interaction for large distances is given. For the case that the charges are in the same atom (small distances) γ_{aa}, the function utilizes the Hubbard parameter U_a[87] which is the difference between the ionization potential and the electron affinity of the atom.

In the DFTB2 method charged and neutral atoms are treated in the same way, because the Hubbard parameter does not depend on the charge fluctuations of the atom, which can vary a lot between charged and neutral atoms. The DFTB3 approach includes in the Hubbard parameter the third derivative of the energy with respect to the charge fluctuation, representing the third order term.

$$U_a^d = \frac{\partial^3 E^{atom}}{\partial q_a^3} \tag{1.9}$$

This term is important for highly localized charges. The energy for the DFTB3 approach is then written as:

$$E_\rho = \sum_i \epsilon_i + E_{rep} + \frac{1}{2}\sum_{ab} \delta q_a \delta q_b \gamma_{ab} + \frac{1}{6} U_a^d \delta q_a^3 \tag{1.10}$$

The Hubbard parameter enhances the proton affinity and the hydrogen bonding energies, which is important for describing organic molecules and especially important for biological molecules, e.g. proteins, where a slight change in the hydrogen bonding energies could validate a mechanism.

4.2 Orthogonalization Model 2 with Multireference Configurational Interaction (OM2/MRCI)

DFTB is a semi-empirical method based on DFT and inherits all the properties of it. Since their are several drawbacks, which do not allow the description of retinal excitations, a semi-empirical wave function based method is required. The OM2/MRCI method offers a balance between computation time and accuracy.

The time consuming step in wave function based methods is the computation of the off-diagonal elements of the Fock matrix. That is why semi-empirical methods, which are based on Hartree-Fock launch at this part of the computationally costly calculation and find ways to neglect these integrals. There are several approximations, which can be made within in a semi-empirical method. Firstly, only valence electrons are considered or the core electrons can be combined and concentrated on the nuclei by reducing nuclear charge or introducing model functions for the combined repulsion. Furthermore, only a minimum basis set can be applied for the valence electrons, like slater type exponential basis sets and at least the overlap matrix S in the Rothaan Hall equation (Equation 2.1) can be set equal to the identity matrix:

$$FC = SC\epsilon \tag{2.1}$$

where F is the Fock matrix, C is the matrix of the LCAO-coefficients, S is the overlap matrix and ϵ describes the diagonal matrix containing the orbital energies.

In the *zero differential overlap* (ZDO) approximation, which is the central assumption for many semi-empirical methods, the overlap between pairs of different orbitals is set to zero (Equation 2.2). Thus, all three- and four-center two electrons integrals are set to zero.

$$\int \phi_\mu \phi_\nu d\nu = 0 \tag{2.2}$$

The differential overlap is called monoatomic differential overlap, when ϕ_μ and ϕ_ν are on the same atom ($\mu = \nu$), and is called diatomic differential overlap, when they are located on different atoms ($\mu \neq \nu$). The remaining integrals are parametrized based on values of *ab initio* calculations or experimental data.

In semi-empirical methods the amount of neglected integrals and parametrization of the remaining integrals varies and gives rise to several different semi-empirical methods: i)

complete neglect of differential overlap (CNDO), ii) the intermediate neglect of differential overlap (INDO) and iii) the neglect of diatomic differential overlap (NDDO). NDDO neglects the differential overlap between atomic orbitals on different atoms. In other words, the method includes the two-electron, two-center integrals $(\mu\nu|\lambda\sigma)$, where μ and ν are located on the same atom A and λ and σ are located on the same atom B. It is the most accurate of the traditional integral approximations.

However, in the established semi-empirical methods the ZDO approximation neglects orthogonalization effects. Thus, in order to correct this missing effect, in the OMx methods the orthogonalization correction is applied on the NDDO method. Furthermore, the *graphic unitary group approach* (GUGA) speeds up the calculations and excitation energies are received in an efficiently manner. Therefore, the active space can be chosen in adequate size, in order to include all relevant orbitals automatically in the calculation.

The excitation energy obtained by OM2/MRCI is usually ≈ 0.3 eV overestimated compared to SORCI values. However, this shift was found to be systematically and comparison of the relative excitation energies is still possible and produces meaningful results.

The short computation time of OM2/MRCI provides an advantage for the simulation of absorption spectra, since excitation energies needs to be calculated on a huge number of snapshots (1000-10000) from a dynamic simulation trajectory.[91]

5 | Force Field Based Methods and Free Energy Methods

5.1 Molecular Mechanics

If processes like ligand binding or protein domain motion reach timescales of several ns up to µs, quantum mechanical methods will be inapplicable, because the detailed description of the electronic structure is computationally very costly. Therefore, further assumptions are necessary to speed-up these calculations. This can be provided by classical molecular mechanics, here the complex chemical bond is reduced to a simple spring-mass model, where the chemical bond is represented by a spring and the nuclei by a mass with a radius. This significantly simplifies the problem and Newtonian mechanics can be applied on the molecules.

Therefore, the inter atomic interactions are defined by a so-called force field, which represents the potential energy of the system and is a function of all atom coordinates. For a complete description of a molecule, internal and external interactions have to be included in the force field. While the internal interactions represent the bonded interactions within a molecule and the external interactions describe Coulomb and Van der Waals interaction between atoms from the same molecule or from different molecules. The atoms are defined in atom-types, dependent on their bond partner in order to distinguish between different bond types and the resulting properties.

The internal interactions consist of three types and describe bonds, angles and torsions angles in the molecule and are usually constituted by harmonic potentials for the bonds and periodic potentials for the angles and the torsions. The parameters are determined empirically and reproduce the behavior of diverse molecules, e.g. the force constants (k in Equation 1.1) or the equilibrium values ϑ^0 or ω^0. The expression of the bonding energy is given in Equation 1.1 where r_{ij} is the bond length between atoms i and j, ϑ_{ijk} the angle between three atoms i, j, k and ω_{ijkl} the torsion angle between the four atoms i, j, k, l.

$$E_{bonded} = \sum_{i,j}^{bonds} \frac{1}{2} k_r^{ij} (r_{ij} - r_{ij}^0)^2 + \sum_{i,j,k}^{angles} \frac{1}{2} k_\vartheta^{ijk} (\vartheta_{ijk} - \vartheta_{ijk}^0)^2 + \sum_{i,j,k,l}^{torsion} \frac{1}{2} k_\omega^{ijkl} \left(1 + \cos[n_{ijkl}\omega_{ijkl} - \omega_{ijkl}^0]\right) \tag{1.1}$$

The Van der Waals interactions are non-polar interactions between two atoms and are described by the external interactions in the force field equation. Two parts are considered in the formula. A long range attractive part, which results from induced dipoles and a short range repulsion part which results from the Pauli-repulsion. These two many-body interactions are combined in the Lennard-Jones potential E_{LJ} and are simplified by two experimentally ascertainable constants ϵ_{ij} and σ_{ij}, being r_{ij} the distance between atom i and j.

$$E_{LJ} = \sum 4\epsilon_{ij} \left(\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma ij}{r_{ij}} \right)^{6} \right) \tag{1.2}$$

Since the nuclei and the electrons are charged, electrostatic interactions, the Coulomb interactions E_{Coul}, occur within the electrons and the nucleus of one atom and between electrons and nuclei of different atoms. The correct description is a challenging part in quantum mechanical methods, especially as every interaction between electrons and nuclei of every atom needs to be considered, the calculation assumes huge proportions and is the most time consuming step. To avoid this, force fields simplify the Coulomb interactions, thereby that the atoms are represented by partial atomic charges q_i and the interaction between the atoms are calculated from the distance r_{ij} between two partial atomic charges and the dielectric constant ϵ_0 (Equation 1.3). However, several quantum mechnical effects are neglected and a description of processes, where electrons are involved is not feasible anymore.

$$E_{Coul} = \sum_{i<j} \frac{1}{4\pi\epsilon_0} \frac{q_i q_j}{r_{ij}} \tag{1.3}$$

5.2 Molecular Dynamics

The force field describes the potential energy of the system and in order to get thermodynamical quantities like kinetic energy, the velocities and the time-variation of the positions and the momenta, need to be included. Therefore, it is necessary to solve Newton's equations of motion for the given potential. This brings up the derivative of the potential energy $E(x_n)$ and represents the force F_n acting on the atoms.

$$F_n = m_n \cdot a_n = -\frac{\partial E(x_n)}{\partial x_n} \tag{2.1}$$

The resulting differential equation has n dimensions and is only possible to solve numerically.

The most commonly used algorithm for this is the Verlet leapfrog algorithm (eq. 2.2).

$$\begin{aligned} x(t+\Delta t) &= x(t) + v(t+\frac{1}{2}\Delta t)\Delta t \\ v(t+\frac{1}{2}\Delta t) &= v(t-\frac{1}{2}\Delta t) + a(t)\Delta t \end{aligned} \tag{2.2}$$

But the initial positions and velocities are needed. The positions of the atoms can be obtained from experimental data like crystal structures, while initial velocities are taken from a Gaussian distribution generated at the temperature chosen for the simulation, based on the Maxwell-Boltzmann distribution. New positions and velocities are then calculated in an alternating fashion.

5.3 Free Energy Calculations

The ergodic theory[92–94] states that the time average of a system $\langle A \rangle_t$ is the same as the ensemble average of the system $\langle A \rangle_e$. This proof is utilized to calculate thermodynamic properties like free energy of a MD simulation.

$$\langle A \rangle_t = \langle A \rangle_e \tag{3.1}$$

MD simulations can sample the correct phase space $\rho(r,p)$ of a system, where r are the coordinates and p are the momenta of the atoms, with a long enough simulation time t, so that all points of the phase space are passed within the trajectory. Then, the thermodynamic ensemble properties like the system energy E can be obtained through the trajectory:

$$\lim_{t\to\infty} \frac{1}{t} \int_{t_0}^{t_0+t} A(t)dt = \int\int A(r,p) \cdot \frac{1}{Q} \cdot exp\left[\frac{-E(r,p)}{k_B T}\right] drdp \tag{3.2}$$

where k_B is the Boltzmann constant,T is the system temperature and Q the partition function of the system. Herewith, thermodynamic quantities are accessible in MD simulations.

The challenge for MD simulations is to sample the correct phase phase, i.e. structural rearrangements of amino acids in proteins, which occur within timescales of several nanoseconds, this is reachable with MD simulations. But if the barrier is larger than several $k_B T$, the required sampling time can be expanded up to µs or even ms. This means, a several nanosecond long simulation at room temperature will not overcome this barrier within the simulation and a longer sampling is required to reach another structural arrangements, which is possibly the global minimum. Therefore, a sufficient sampling is crucial for the proper description of the thermodynamic properties of the system, since an incorrectly sampled conformation in the simulation leads to wrong conclusions.

Another problem arises, when the minima are known, but the thermodynamic properties like free energy of the barrier $\Delta G^{\dagger}$ needs to be determined. The MD simulation will not sample

often enough the higher lying conformations with too short simulation time. However, for the correct estimation of the free energy of the barrier $\Delta G^{\dagger}$, the higher lying conformations are crucial and contribute significantly to the free energy barrier. Therefore, a too short simulation time or a insufficient sampling increases the error of $\Delta G^{\dagger}$ drastically or makes the determination of $\Delta G^{\dagger}$ impossible.

To overcome these problems, new techniques are required to negotiate the sampling problem and to provide a correct prediction of the properties of a molecular system. These methods are called enhanced sampling methods.

5.3.1 Metadynamics

With Metadynamics,[95] the free energy landscape of a system is explored and reconstructed using an external, history-dependent bias potential $V(\xi, t)$. The potential is defined by the collective variables $\xi_i(r)$ and is added to the Hamiltonian of the system. The system energy will be biased by this potential. The conformational minima will be filled up and thus, the system will be pushed out of a minimum conformation. Usually the bias potential is a sum of Gaussian potentials, where σ_i is the Gaussian width and ω the Gaussian deposit rate:

$$V(\xi, t) = \int_0^t dt' \omega exp \left[- \sum_{i=1}^{d} \frac{(\xi_i(r) - \xi_i(r(t')))^2}{2\sigma_i^2} \right] \tag{3.3}$$

More in detail, already sampled conformations along the collective variable get energetically penalized by the added Gaussian potential and it gets easier for the system to reach other regions of the phase space. Therefore, the system can overcome barriers and reach new energetic minima (see Figure 5.3.1).

In the well-tempered metadynamics, the Gaussian height is adjusted during the simulations and decreases with proceeding simulation time. The initial Gaussian height ω_0 gets decreased dependent on the deposit frequency τ and the time-dependent potential height ω is calculated by:

$$\omega = \omega_0 \tau exp \left[- \frac{V(\xi, t)}{k_B T} \right] \tag{3.4}$$

Whereby the parameters σ_i, ω_0 and τ control resolution of free energy profile and have to be defined at the beginning of the simulation.

5.4 Quantum Mechanics / Molecular Mechanics approach

The simulation of large biomolecules by classical force fields arises some problems especially for the description of photo-active proteins. The process which takes place

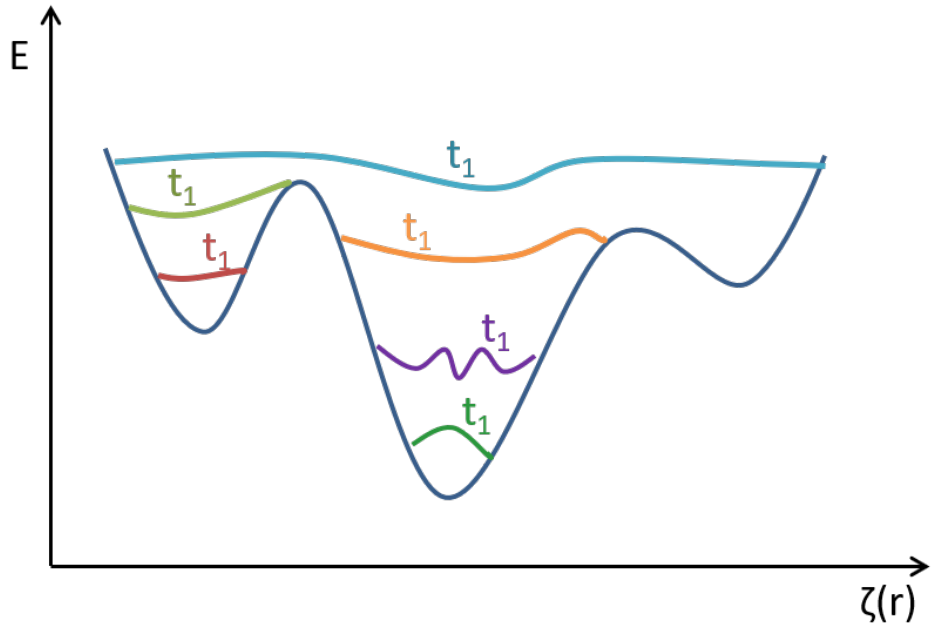


Figure 5.3.1: Enhanced sampling of a free energy profile with well-tempered metadynamics. Gaussian potentials are added dependent on the already sampled phase space during the simulation and therefore enable an enhanced sampling of further areas of the phase space.

during photo activation has to involve changes and rearrangements of the electron density. Molecular mechanical force fields are not able to describe these phenomena, because they do not consider explicitly electrons. All processes where electrons play a role can not be described by force fields, e.g. charge transfer, chemical reactions or photochemical processes. These processes have to be described by a quantum mechanical method. For large systems, e.g. proteins, it is impossible to treat the whole system quantum mechanically, because of the infeasible computational cost. Usually the change of the electronic structure affects only on a small region of a system and is therefore enough to treat this small region with quantum mechanical methods and describe the rest with a classical force field. The interaction between both QM and MM subsystems has also to be taken into account. The total energy E_{tot} of the system is given as:

$$E_{tot} = E_{QM} + E_{MM} + E_{QM/MM} \tag{4.1}$$

QM and MM subsystems can be coupled at diverse levels. The most simple way is *mechanical embedding*. The interactions between QM and MM regions are described with the classical Lennard-Jones potential (eq. 4.2). The charges for the QM region are given by Mulliken charges q_i^{Mull} and the vdW parameters are taken from the force field.

$$E_{QM/MM} = \sum_{i}^{QMatoms} \sum_{m}^{MMatoms} \left(\frac{q_i^{Mull} \cdot q_m}{r_{im}} + 4\epsilon_{im} \left[\left(\frac{\sigma_{im}}{r_{im}} \right)^{12} - \left(\frac{\sigma_{im}}{r_{im}} \right)^{6} \right] \right) \tag{4.2}$$

The disadvantage of this method is that the Coulomb interactions are underestimated, the

polarization of the wave function through the MM atoms is missing and the charges of the QM atoms will not be updated during the calculation.

A better choice is the *electronic embedding*, where the point charges of the MM atoms are taken into account in the Hamiltonian of the QM calculation. Thus a polarization of the electronic density by the MM charges is ensured. This is achieved by an additional term in the Hamiltonian:

$$\hat{H}^{el}_{QM/MM} = \sum_{i}^{QMelec} \sum_{m}^{MMatoms} \frac{q_m}{|r_i - r_m|} + \sum_{I}^{QMnuc} \sum_{m}^{MMatoms} \frac{q_m Z_I}{|R_I - r_m|} \tag{4.3}$$

where q_m is the point charge of the MM atoms and Z_I is the nuclear charge of the QM atoms. In this embedding scheme the QM region is affected by the MM region, but the MM region notices no influence on the QM region. To achieve a MM polarization, a polarizable force field has to be considered that reacts to the changes of the electron density. This is only necessary when high accuracy is needed, being aware of the high computational cost required.

The afore mentioned embedding schemes consider only non-bonded interactions. Often the selection of the region includes a covalent bond crossing the boundary between the QM and the MM region, especially in proteins where only some amino acids are treated with quantum mechanics and the rest with classical force fields. The boundary in proteins are usually set between C_α and C_β atoms. The simple cutting of the bond will give artifacts, since the QM-region is not saturated, leading to unrealistic descriptions. In order to ensure that the QM region is a complete molecule, a link atom (LA) is placed between C_α and C_β atoms. To avoid interactions and overpolarization at the border between QM and MM region the charges of the nearby MM atoms are deleted and smeared on the next atoms.

For QM/MM calculations it is important to choose the border carefully, in order to avoid that the QM region is unnecessary too large due to computational cost or that the QM region is too small, possibly giving rise to artifacts in the QM region.

Part III

Results

6 | Dark State of Channelrhodopsin-2

In the photocycle of Channelrhodopsins successive processes like proton transfer reactions, reorientation of single amino acids, rearrangements of the protein backbone and consequential helix movements are crucial for the opening and closing of the ion channel. Structural adjustments in rhodopsins during the photocycle usually occur due to energetic differences of only a few kcal/mol between two states. Thus, small variations in the initial structure pattern of the protein could result in an incorrect description of the subsequent states. Therefore, understanding and investigating the structure-function relationship and the mechanisms occuring during the photocycle requires a detailed and precise knowledge of the hydrogen bond pattern, the amino acid arrangements and their characteristics in the initial dark-state structure. The topic of this chapter is the detailed characterization of the amino acid side chain orientation, the retinal conformation and the hydrogen bonding pattern in the active site and in the bonding pocket of Channelrhodopsin-2 (ChR-2).

Several spectroscopic studies concerning the investigation of the photocycle had been done in the past providing a contradictory picture of the retinal conformation, the amino acid protonation state and the rearrangement in the amino acid side chains during the photocycle (details see section 2.1). This information serves as a basis for investigation of the initial dark-state structural pattern of the ChR-2 photocycle.

The missing crystal structure and the inaccuracy of a generated homology model limits the study to the retinal and the amino acids surrounding the binding pocket. Spectroscopic measurements support the results and conclusions of this theoretical study. The theoretical work presented here, will give an insight in atomistic description of the active site and can clarify the structural dependencies in the binding pocket, which lead to a deeper understanding of the complex mechanisms in ChR-2.

6.1 Absorption Spectrum of Channelrhodopsin

The absorption of light by the retinal chromophore is the initial event in the rhodopsin photocycle. Since the absorption spectrum displays the influence of the environment on the retinal, a detailed inspection of the spectrum offers first information about the binding pocket

in Channelrhodopsin-2. Given that, the retinal conformation, such as its planarity and bond length alternation (BLA) as well as the charge distribution and the hydrogen bond pattern of the Schiff base ($RSBH^+$) determine the absorption property of the retinal. These structural factors and electrostatic interactions are tuned by the arrangement of the surrounding binding pocket in the rhodopsins.

The global absorption maximum in ChR-2 (see Figure 6.1.1) is at 473 nm (2.62 eV)[48] and is blue shifted about 95 nm with respect to BR (568 nm, 2.18 eV).[96,97] Former studies have shown that the blue shift between the two rhodopsins is mainly caused by the different amino acid compositions in the binding pocket.[98,99] Furthermore, the ChR2 spectrum forms two additional sub-peaks at 442 nm (2.81 eV) and 414 nm (3.00 eV). In contrast, the BR spectrum shows only a single peak. In the BR absorption spectrum the single peak comes from the known structural arrangement in the vicinity of the retinal $RSBH^+$, which is organized in a static so-called pentagonal cluster of the two counter ions and three water molecules. Since Lórenz-Fonfría *et al.*[60] reported that the absorption of the retinal chromophore in ChR-2 consists exclusively of $S_0 \rightarrow S_1$ electronic transitions, these two sub-peaks in the absorption spectrum are either due to excitation in various vibrational levels or could come from a heterogeneous active site structure based on the one structure - one peak relation in BR.

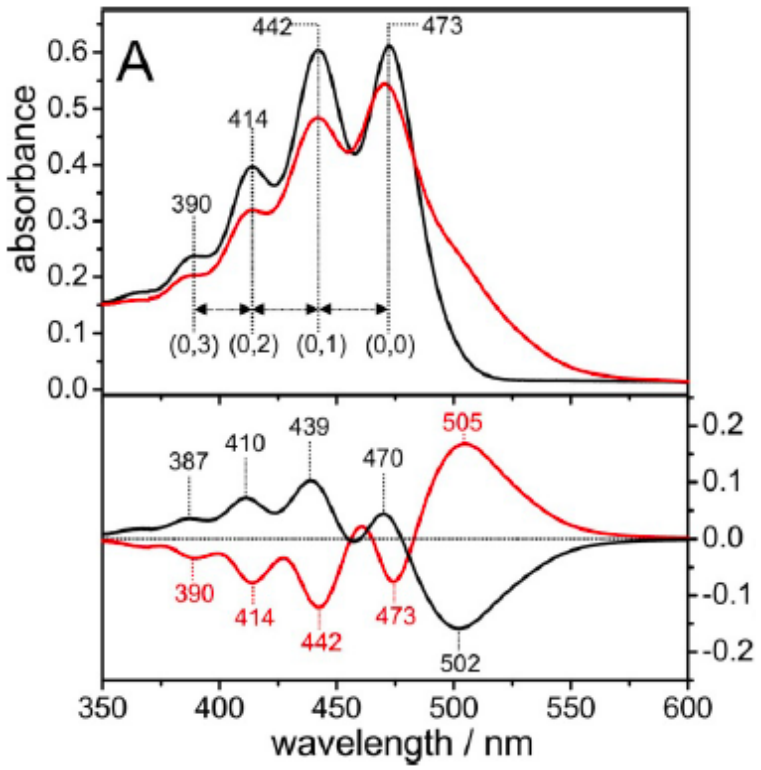


Figure 6.1.1: Experimental absorption spectrum of ChR-2-H134R mutant at 4 K (upper panel, black: before illumination at 455 nm (2.7 eV), red: after illumination at 455 nm) and the resolved fine structure (second derivative spectrum) in the lower panel adapted from Bruun *et al.*[48]

6.2 Dark-State Active Site and DC-Gate

After the initial absorption of light, the retinal isomerizes in the excited state and initiates the photocycle of the ChR-2. During the photocycle the pore-opening of the channel is initialized by a proton transfer from the $RSBH^+$ of the retinal to one of the counter ions E123 or D253.[39,100] It is still under debate which of the two counter ions act as the proton acceptor.[32,39,47,56,64] Another crucial structural pattern for understanding the mechanisms in the early states of the photocycle is the detailed active site hydrogen bond network in the initial dark-state, before the isomerization. But even this structural pattern is not yet fully elucidated.[42,43] One of the main issues in this debate is the $RSBH^+$ hydrogen bond pattern and the water distribution in the vicinity of the $RSBH^+$. In the crystal structure of the related C1C2 chimera, only one water molecule is present next to the $RSBH^+$ and a direct salt-bridge to the E123 counter ion is formed.[30] In contrast, resonance Raman spectroscopy[43,48] and theoretical studies[42] indicate a water bridged hydrogen bond of the $RSBH^+$.

Furthermore, the extended hydrogen bond network in the binding pocket, especially the hydrogen bonding situation for the residues D156 and C128 (DC-Gate), which are involved in the photocycle,[56,59] is not completely solved. However, D156 and C128 are crucial for the function mechanism of the ChR-2, thus mutation studies on these residues extend the open-state lifetime about 200- to 1000-fold compared to the ChR-2 wild type. Spectroscopic measurements suggest a direct hydrogen bond between the D156 and C128 residues,[39,49,64,101] while MD simulations support a water bridged DC-Gate.[42,102,103]

From spectroscopic experiments[47] it is known, that the E90 residue located in the central gate deprotonates during the photocycle[34,47,52,55,58,64] and a downward flip of the side chain towards the counter ion E123 has been observed in MD simulations by Kuhne *et al.*[47] But this mechanism is not yet affirmed.

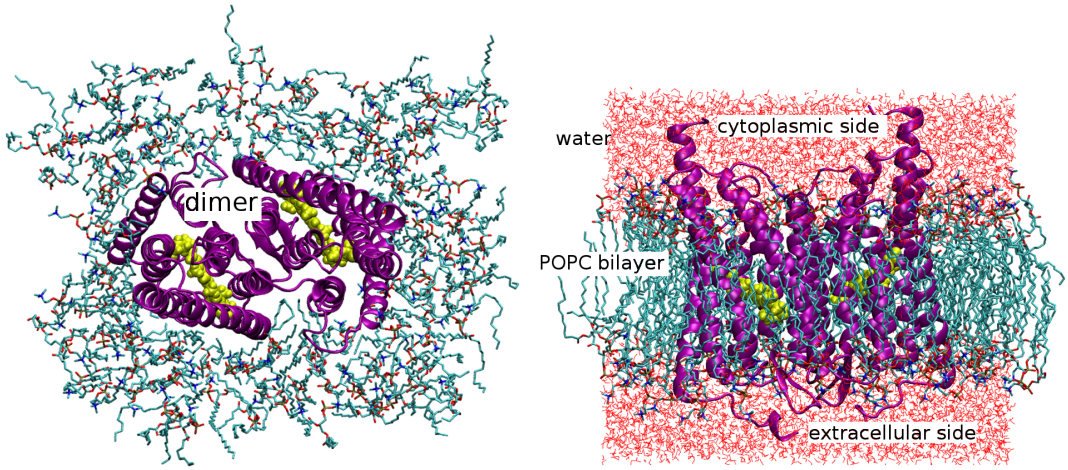
The retinal conformation in the dark-state is still a controversially discussed subject. Resonance Raman spectroscopy experiments[43] at 298 K as well as the FT-IR difference spectroscopy experiments[49] at room temperature detected two bands, which indicate a mixture of all-*trans* and 13-*cis*,15-*syn* retinal in the protein. Bruun *et al.*[48] found an initial dark adapted state (IDA) with a pure all-*trans* conformation while a minor 13-*cis*,15-*syn* conformation appears after first illumination of the protein. Controversially, solid state NMR experiments[50] at 100K suggest a pure all-*trans* conformation in the binding pocket after several hours in darkness.

6.3 Computational Details

6.3.1 Rhodopsin Models

A crystal structure of ChR-2 has only been available since November of 2017, which made the use of a homology model based on the C1C2 chimera unavoidable for the elucidation of the functional mechanisms of ChR-2 on the molecular level in the course of this work. In former studies, homology models based on *Anabaena* sensory rhodopsin and the C1C2 chimera had been studied[37,42] regarding the water distribution, the hydrogen bonding pattern of the $RSBH^+$ and intrahelical hydrogen bonds.

A homolgy model based on the C1C2 chimera (PDB-Code: 3UG9[30]) from Watanabe[42] is adopted and additional water molecules are incorporated in the protein structure using the DOWSER package.[104] The positions of the water molecules are determined by assessing the hydrophilicity of an existing cavity. The ChR-2 is duplicated and constructed as dimer[105] and embedded in an equilibrated POPC (1-Palmitoyl-2-oleoylphosphatidylcholine)-bilayer with a surrounding water box as bulk solvent (see Figure 6.3.1).



(a) Top View (b) Front View

Figure 6.3.1: Model of ChR2 dimer (violet) embedded in a POPC-bilayer (cyan) and surrounding water molecules (red).

As discussed above, the hydrogen bonding pattern in the active site is not clear. Previous studies with standard force field methods indicate a salt bridge between the $RSBH^+$ and the two counter ions.[37,42] In general, classical force fields overestimate electrostatic effects and underestimate electronic effects since they neglect the explicit description of electrons, including only a partial charge on each atom. This leads to an overestimation of the attraction between atoms with opposite charge and to an underestimation of the hydrogen bond strength. Therefore, classical force fields can lead to a wrong description of the

interactions within the active site. This suggestion was already proposed for the binding pocket of BR,[6] indicating that quantum mechanical methods have to be taken into account in order to describe correctly the electronic interactions. Because of the high computational cost of QM methods, the simulation time which could be achieved is not long enough to observe a rearrangement in the structure, which takes place in the range of nanoseconds to milliseconds. Therefore the strategy applied in this work is to model the active site by proposing initial guesses: a water molecule is placed next to the $RSBH^+$ avoiding the formation of a salt bridge with E123. This assumption required an equilibration of the surrounding residues, keeping the replaced water molecule and the side chain of E123 in a fixed position. To test this, the structure of the active site is modeled by moving the side chain of E123 towards T127 and moving a nearby water molecule next to the $RSBH^+$ (see Figure 6.3.2) and a hydrogen bond to that water molecule is fixed by distance restraints at a distnace of 0.18 nm. The two oxygen atoms from the side chain of E123 are fixed by harmonic position restraints with a force constant of 10000 $kJ\,mol^{-1}\,nm^{-1}$ during the equilibration procedure.

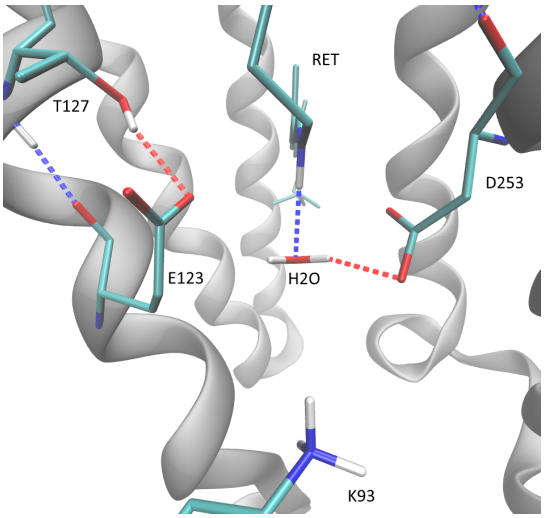


Figure 6.3.2: Modified binding pocket of ChR2. A water molecule is moved nearby the $RSBH^+$ and E123 is moved towards T127. The hydrogen bond between T127 and E123 side chain is constrained.

For titrable residues, standard protonation states are assumed, except D156 and E90, which have been modeled as protonated.[30,42] E90 is known to undergo deprotonation during the photocycle[34] and D156 has been determined as protonated in spectroscopic measurements.[34,37,60]

6.3.2 MM Equilibration

The setup of the simulation box and the active site modeling as well as the equilibration procedure was part of the master thesis of the author and the detailed equilibration procedure is described there.[106] In this part only important adjustments are repeated.

For the MM part the CHARMM36[107,108] force field is used with additional parameters to describe the retinal chromophore. The TIP3P model[109] is applied to describe water molecules. All MD-simulations were performed with the GROMACS package version 4.5.5.[110,111] A step wise equilibration of the bulk solvent, the POPC-bilayer and the protein was performed, while the active site (retinal, E122, D253, hydrogen bonded water molecule) of the protein was frozen, in order to receive the structural adjustments.

However, during equilibration, the water bridging the DC-Gate in monomer 2 was lost, while it remains in monomer 1. Thus, two DC-Gate structure motifs have been found one with the water-bridged DC-Gate and one with the direct DC-Gate, where a direct hydrogen bond between the C128 and D156 was formed. The starting structures of both monomers before the equilibration contained the water as shown in Figure 6.4.8a. Since this motif have been controversially discussed in former studies, the two DC-Gate structures have been validated by stretching vibrational frequency calculations of D156 side chain (νCOO) using the normal mode analysis (NMA) with the CHARMM37b1 program at the DFTB3/MM level of theory. This method has been validated in detail in the study by Welke *et al.*[102] The residues C128, D156, and the bridging water molecules (for water bridged DC gate) were included in the QM region, and the rest was treated as MM part. The calculated νCOO value was 1730 cm−1 for the water bridged DC gate structure, which agrees well with the experimental data (1735 cm^{-1}),[101] and 1693 cm−1 for the direct DC gate structure, i.e. 42 cm−1 red shifted compared with the FTIR value. These frequency calculations have been taken as a strong indication for the presence of one water molecule in the DC gate. Thus, the monomer 2 structure is discarded in the following analysis.

6.3.3 QM/MM Simulation

MM force fields fail to describe the hydrogen bonding pattern of rhodopsins correctly,[112] especially in MM methods the salt bridge bond is energetically favored, thus the pentagonal hydrogen bond arrangement in BR breaks, when using a MM force field.[42,113,114] Therefore, QM/MM methods are required to characterize the active site. In the following procedure the QM-zone contained the retinal chromophore with the covalently linked K257, the side chains E123, D253, K93, and three water molecules.

Since the modified active site was not adapted to the surrounding protein structure and this could lead to convergence problems with the QM/MM interface, the restraints on the active site have been removed in a step wise smoothed relaxing procedure (see Table 6.3.1).

50 structures out of the simulation in the last step are chosen and a QM/MM minimization is performed in order to get the minimized energy. In the 50 optimized geometries several hydrogen bond patterns are found. This means the ChR-2 structure has no well-defined

Table 6.3.1: Procedure for reducing the position restraints in the QM/MM-simulations. The procedure takes a total time of around 7 weeks for one structure.

step	simulation period [ps]	position restraints on atoms [kJ mol^{-1} nm^{-1}]	distance restraints of H-bond	QM-water position restraints [kJ mol^{-1} nm^{-1}]
step1	200	5000	yes	500
step2	200	3000	yes	500
step3	200	2000	yes	500
step4	200	1000	yes	500
step5	200	500	yes	500
step6	200	0	no	500
step7	600	0	no	500

ground state minimum like in BR. Up to here, the model building and the extended equilibration was part of the master thesis of the author and is published with Guo *et al.*[115]

Because of this high variety of hydrogen bond patterns in the optimized structures, extended QM/MM simulations in order to improve the sampling and allow investigation of the active site motifs are performed.

A 1 ns extended QM/MM simulation was performed using the structure after the simulation of step7 as starting structure. Since in this extended simulation, the hydrogen bond between E123 side chain and T127 side chain breaks and a downward movement of the E123 side chain has been observed, the T127 side chain was added to the QM-zone. Thus, an accurate description of the stability of the hydrogen bond between E123 and T127 side chain can be ensured and a bias caused due to the wrong description of the force field is avoided. Then seven randomly selected structures from the extended 1 ns simulation have been chosen as starting structures and seven 1 ns simulations have been performed. In the production runs, the QM-zone consists of the retinal with the covalently linked K257, the side chains E123, D253, K93 and T127 and four water molecules and is called large-QM-zone. The temperature was set to 300 K.

13-*cis*, 15-*syn*

Concerning the debate, whether pure all-*trans* retinal conformation or a mixture of all-*trans* and 13-*cis*, 15-*syn* retinal conformation is present in the dark-state, the 13-*cis*, 15-*syn* retinal from BR (PDB code: 1XIO) has been aligned to the all-*trans* retinal. Afterwards, the all-*trans* has been replaced in the ChR-2 model by 13-*cis*, 15-*syn* retinal. For this, six randomly selected structures of the all-*trans* trajectories with large-QM-zone have been used. The new generated ChR-2 models have been minimized with the MM method and position

restraints were placed on the active site, followed by a QM/MM equilibration step with the large-QM-zone. The structures have been used as starting structures for the production QM/MM-MD runs containing 13-*cis*, 15-*syn* retinal.

Low Temperature Simulations

Since experiments at low temperatures[50] and former studies on BR revealed a strong temperature dependency of the $RSBH^+$ orientation,[113] QM/MM simulations at 100 K have been performed. 14 QM/MM trajectories each with a time scale of 1 ns and a temperature decrease from 300 K to 100 K within 10 ps have been performed. The temperature was kept at 100 K for the remaining simulation time. 14 snapshots, carrying the three hydrogen bonding patterns and two E123 side chain conformations, of the production QM/MM simulations have been randomly chosen as starting structure.

Bacteriorhodopsin Simulations

With the same setup as used for the ChR-2 QM/MM MD simulation a 1 ns BR trajectory on the BR structure from a previous study[99] has been performed. Included in the QM-zone are the retinal chromophore with the covalently bound K219, the D212 and D85 counter ions and R82 side chains and the three water molecules which form the pentagonal cluster structure.

Simulations on Crystal Structure

The computational study of the ChR-2 has been performed on the homology model based on the C1C2 chimera. In late 2017, a crystal structure of ChR-2 has been published,[31] therefore additional studies have been performed on this newly published structure. The crystal structure has been aligned on the position of randomly chosen snapshots of the existing ChR-2 models and the homology has been replaced by the crystal structure. Different simulation setups have been composed:

i) the water molecules of the crystal structure have been replaced by the water molecules of the study on the homology model,

ii) the water molecules of the crystal structure have been replaced by the water molecules of the study on the homology model but the water molecule next to the $RSBH^+$ has been deleted

iii) only the water molecules of the crystal structure have been used.

Furthermore, to investigate the stability of the E90 and the correlation between the DC-Gate and the active site, the amino acids which are included in the QM-zone are varied for each snapshot. The four QM-zone setups consisting of:

a) retinal, D253, E123, T127, K93 (large-QM-zone)

b) retinal, D253, E123, T127, K93, E90 (QM-zone-e90)

c) retinal, D253, E123, T127, K93, C128, D156 (QM-zone-dcgate)

d) retinal, D253, E123, T127, K93, C128, D156, E90 (QM-zone-dcgate-e90)

For setup i) three, with setup ii) and ii) respectively two snapshots have been chosen and QM/MM simulations with each of these snapshots have been performed. The QM-zone has been varied as described in point a) to d) in all QM/MM simulations. At the end 28 trajectories have been simulated with varied water clusters at the vicinity of the $RSBH^+$ and varied QM-zones. Each trajectory has a timescale of 1 ns.

All QM/MM-simulations in this study have been carried out with the GROMACS package version 5.0.4[110,111]. For the MM part the CHARMM36[107,108] force field has been used, while the SCC-DFTB3 method[88,116,117] with the 3ob-parameter set[118] was applied to the QM region. The QM/MM boundary was chosen between C_α and C_β atoms of the amino acid side chain and the standard link atom approach was used. The time step was set to 1 fs.

For reliable modeling of the ChR-2 active site it is essential to describe hydrogen bonded networks with a high accuracy as in full DFT calculations using medium-sized basis sets, DFTB3/3ob has shown to fulfill these requirements.[88] The DFTB3/3ob method is preferred for the calculations on ChR-2, because the required simulation time for such flexible systems lies in the range of nanoseconds and with the computationally more expensive DFT method it is not possible to sample in these large timescales. It was shown in the last few years that, when the DFTB/MM method is applied to infrared and absorption spectroscopy it is sensitive to the active site structure.[102,119,120]

6.3.4 Excitation Energies

Reproducing the experimental absorption spectrum is an indication of a good active site structure, since the ChR2 structure does not have a well-defined minimum in the dark-state, the excitation energies need to be calculated on a statistically meaningful set of geometries. Thus, the excitation energy along the trajectories has been calculated so that 13 000 excitation energies are computed, in order to evaluate the dynamical effects of the flexible structure on the absorption.

For the calculations of the excitation energies the OM2/MRCI method has been chosen. Since the method has a semi-empirical approach, the method needs to be evaluated by a *ab initio* method. It has to be tested if OM2/MRCI is able to reproduce the color shift due to geometrical variations of the retinal and the electrostatic interactions of the surrounding amino acids. The benchmark has been performed by Yanan Guo and the details are reported in her PhD-Thesis[121] The OM2/MRCI method was found to give a systematic blue shift of

about 0.3 eV compared to the SORCI results and by neglecting protein polarization effects in the energy calculation the SORCI calculations are about 0.1 eV blue shifted compared to the experimental absorption maximum.

For the OM2/MRCI calculations the MNDO program package[122] has been used. The OM2/MRCI calculations employed closed-shell molecular orbitals (MOs) and an active space of 20 electrons in 20 MOs (i.e. the 10 highest occupied MOs and the 10 lowest unoccupied MOs). The reference configurations comprised the closed-shell ground-state configuration and four configurations generated by excitations from the two highest occupied MOs (π, $\pi-1$) to the two lowest unoccupied MOs (π^*, π^*+1), i.e. the single excitations (π) $\rightarrow$ (π^*) and ($\pi-1$) $\rightarrow$ (π^*+1) as well as the double excitations (π, π)$\rightarrow$(π^*, π^*) and (π, $\pi-1$)$\rightarrow$(π^*, π^*+1). All single and double excitations from the reference configurations were included in the MRCI calculations. The same MRCI treatment was applied in the single-point calculations at the snapshot geometries extracted from the QM/MM MD simulations

6.3.5 Free Energy Calculations

The transition between the different E123 side chain conformations have been studied by well-tempered metadynamics.[123–125] Both retinal conformations (all-*trans* and 13-*cis*, 15-*syn*) have been investigated in a 100 ns classical MM simulation. Moreover, the side chain orientation of the E90 side chain was studied by well-tempered metadynamics as well. As the collective variable (i.e. the reaction coordinate), the dihedral angle around the Cβ-Cγ bond was used. The parameters were set as follows: Gaussian height of 0.2 kcal/mol, Gaussian width of 0.2 rad, Gaussian depositing frequency of 2 ps (1000 time steps), bias factor of 4, and temperature of 300 K. The convergence of the metadynamics simulation was evaluated by comparing the free energy as a function of the collective variable at different times and the similar profiles obtained after 85 ns indicates a good convergence of the simulation.

6.4 Results and Discussion

6.4.1 Active Site Structural Motifs

In the 13 QM/MM trajectories (seven trajectories consisting of all-*trans* retinal and six trajectories consiting of 13-*cis*, 15-*syn* retinal) a very flexible and heterogeneous active site has been observed. These results are similar to the QM/MM simulations on the C128T-Mutant[121] and are in contrast to former classical force field studies, where only the salt-bridge hydrogen bond pattern is preferred.[42] Three RSBH$^+$ hydrogen bonding pattern as well as two E123 side chain conformations have been observed (see Figure 6.4.1) in the trajectories.

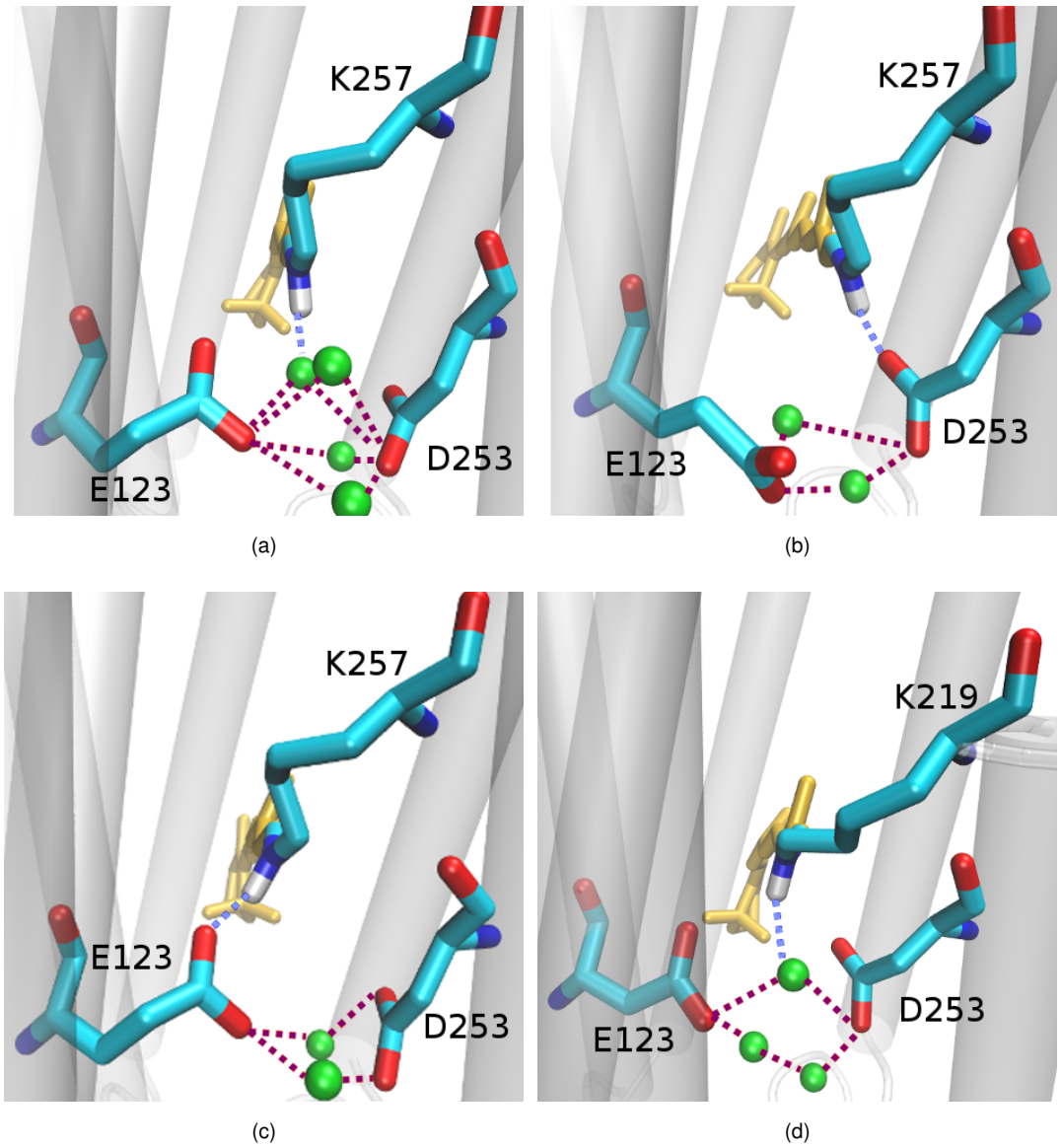


Figure 6.4.1: Hydrogen bonding pattern of ChR2-WT sampled in 14 ns QM/MM simulation. a) $RSBH^{+}...OH_2$ b) $RSBH^{+}...O^{-}$-D253 c) $RSBH^{+}...{}^{-}O$-E123 d) for comparison, the rigid pentagonal cluster of active site of BR.

6.4.2 RSBH$^+$ Hydrogen Bonding Patterns

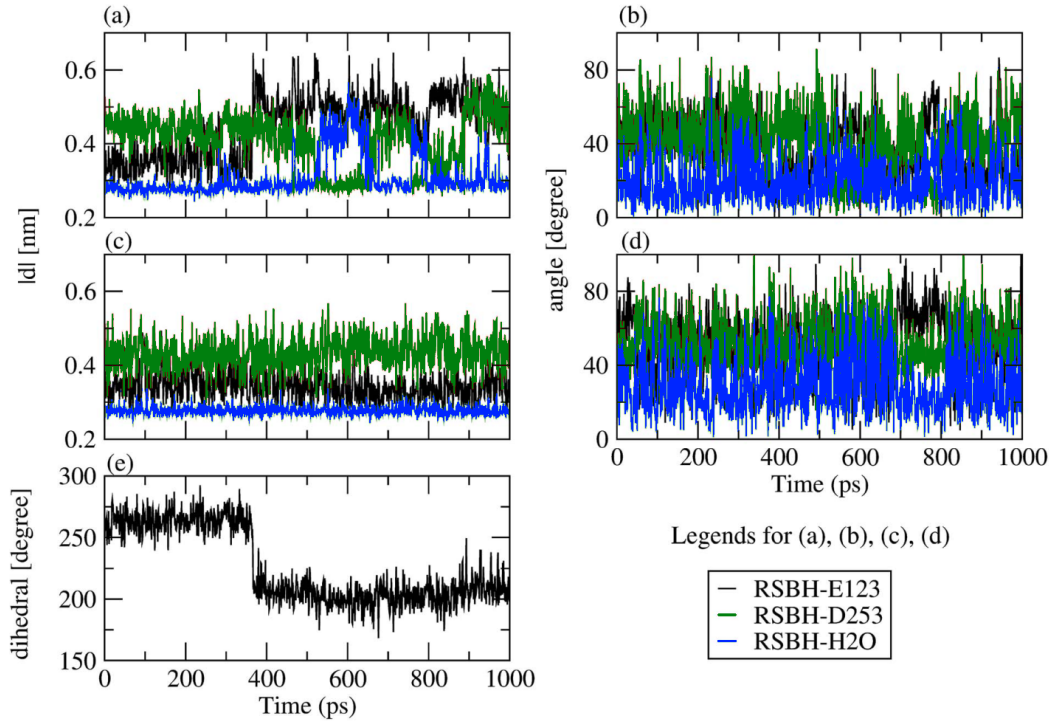


Figure 6.4.2: a)-d) Hydrogen bonding pattern of RSBH$^+$ over 1 ns QM/MM trajectory. a) Hydrogen donor distance and b) Hydrogen donor-acceptor angle of ChR2. c) Hydrogen donor distance and d) Hydrogen donor-acceptor angle of BR. In e) the dihedral angle of the E123 side chain is shown over 1 ns simulation time. The angle smaller than 240° corresponding to the E123-downward orientation of the side chain.

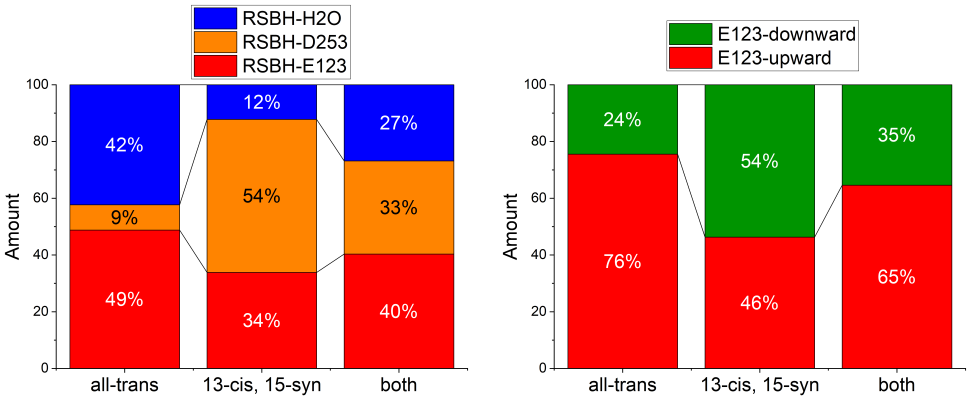
The evaluation of the flexible hydrogen bond pattern in the active site needs adequate criteria to determine the presence of a hydrogen bond. Therefore, in this study the hydrogen bond is defined by the angle between the formal proton H$^+$ acceptor A ($-NH^+...A-$) and donor N. This angle should be bigger than 150° for a hydrogen bond. The second criterium is the distance between the formal proton H$^+$ acceptor A and the donor N and should not be larger than 3.5 Å. Applying these criteria to the 13 trajectories three hydrogen bond patterns of the RSBH$^+$ i) with D253 (Figure 6.4.1b), ii) E123 (Figure 6.4.1c) and iii) with a nearby water molecule (Figure 6.4.1a) are found in the simulations. During the 1 ns trajectories transition between the hydrogen bonding motifs are possible, which is shown in the temporal evolution of the distance and angle criteria in Figure 6.4.2. The clustering of the snapshots is based on the above mentioned criteria. The snapshots, where no hydrogen bond could be found, have been allocated to the hydrogen bonding motif dependent on orientation of the RSBH$^+$, which means this snapshots are clustered dependent on the angle of $-NH^+...A-$.

The trajectories containing 13-*cis*, 15-*syn* retinal show the same flexible behavior as the all-*trans* trajectories and the analysis of their hydrogen bonding pattern brings up the same

structural motifs as in the all-*trans* trajectories. This is not unexpected, since 13-*cis*, 15-*syn* and all-*trans* retinal have very similar structural orientation in the active site of ChR-2 (see Figure 1 in the Appendix).

The percentage of each hydrogen bond pattern in the all-*trans* trajectories (see Figure 6.4.3) shows a preference of the $RSBH^+$...$^-$O-E123 and $RSBH^+$...OH_2 pattern, while in the 13-*cis*, 15-*syn* trajectories the $RSBH^+$...$^-$O-D253 pattern is dominant. Considering both retinal configurations all three patterns are equally represented. This shows that D253 and E123 side chains are both the hydrogen bonding partner of the $RSBH^+$ and therefore could both be the proton acceptor of the $RSBH^+$ proton in the later state of the photocycle.

When comparing the hydrogen bonding pattern with the hydrogen bond network in the BR trajectory, the pentagonal water arrangement is well-preserved in the QM/MM simulation (see Figure 6.4.2 c) and d)), while in ChR-2 the water arrangement is flexible. In contrast, in ChR-2 two water molecules bridging E123 and D253 and two more water molecules are involved in the water cluster when the $RSBH^+$...OH_2 motif is present (see figs. 6.4.1a to 6.4.1c). When two water molecules are involved in the hydrogen bonding network, it is named half barrel and if all four water molecules are involved it will be called barrel.



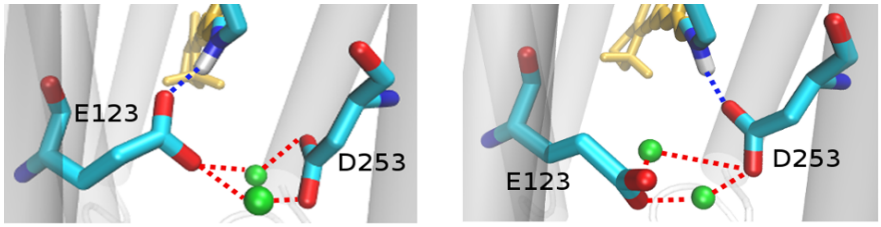
(a) Percerntage of Hydrogen Bonding Pattern of $RSBH^+$

(b) Percentage of E123 Conformation

Figure 6.4.3: Percentage of a) Hydrogen bonding pattern of $RSBH^+$ and b) E123 side chain conformation of all 13 QM/MM Trajectories, broken down into the retinal conformation.

6.4.3 E123 Side Chain Conformation

Two side chain conformations of E123 were sampled in the 13 trajectories: i) a E123 side chain orientation towards the cytoplasmic side (E123-upwards) ii) and a E123 side chain orientation towards extracellular site (E123-downwards) (see Figure 6.4.4). The E123 side chain conformation is characterized by the dihedral angle around C_β and C_γ bond (E123 upward: dihedral angle $>240°$; E123 downward dihedral angle $<240°$). The orientation of the E123 side chain influences the hydrogen bonding pattern at the $RSBH^+$ (Figure 6.4.3),



(a) (b)

Figure 6.4.4: E123 side chain conformation in a) upward conformation towards the cytoplasmic side and in b) E123 downward conformation towards the extracellular side.

when the E123 side chain is upward oriented, the $RSBH^+...OH_2$ and $RSBH^+...^-O$-E123 are preferred, on the other hand when E123 is downward oriented the $RSBH^+...^-O$-D253 is dominant.

To further analyze the E123 side chain movement, well-tempered metadynamic simulation with classical MD have been performed. The free energy profiles of the two E123 conformations have been evaluated and are shown in Figure 6.4.5. In ChR-2 in all-*trans* trajectory, a 2.5 kcal/mol rise in free energy for the transition from E123-upward to E123-downward orientation, indicates a slightly higher stability of E123-upward orientation with all-*trans* retinal in the binding pocket. On the other hand, for the 13-*cis*, 15-*syn* retinal no free energy barrier is observable. In this case there is no barrier between the two conformations, and they have almost the same free energy. This is in good agreement with the QM/MM simulations, since in the all-*trans* trajectories the E123 upward orientation is more prominent, while in 13-*cis*, 15-*syn* the downward orientation is preferred (see Figure 6.4.3).

Concluding this section, it could be said, that the active site is very heterogeneous and flexible compared to BR. Regarding the characteristic absorption spectrum of ChR-2, the E123 side chain conformation and three $RSBH^+$ hydrogen bonding patterns could be responsible for the multi-peaked absorption spectrum of ChR-2. The flexibility of the $RSBH^+$ hydrogen bond induces, that both counter ions could act as proton acceptor.

6.4.4 Low Temperature Simulations and Retinal Configuration

In solid-state NMR studies at of Becker *et al.* at 80 K,[50] the authors observed a shorter distance between RSBH+ and the counter ions in ChR2 than in BR, which is in conflict with the distance of the QM/MM trajectories at 300 K. BR (3.83 Å) and ChR2 (3.82 Å) have almost the same $RSBH^+$ to counter ion distance. Nevertheless, a previous study on BR showed a structural change at low temperatures.[113] To get an insight into the temperature dependence of our system, simulations at 100 K have been performed.

The hydrogen bond pattern remains the same at 100 K, but due to the low temperature

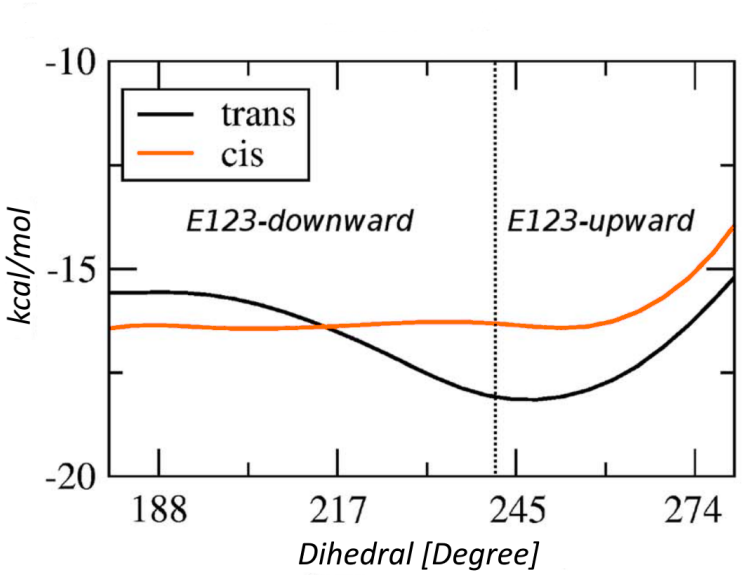


Figure 6.4.5: Free energy scan of E123 side chain rotation. In black the free energy curve of ChR-2 containing all-*trans* retinal and in orange containing 13-*cis*, 15-*syn* retinal

and the decreased kinetic energy of the system, the simulations show almost no transitions between the single hydrogen bonding patterns, while at 300 K several transitions happen within 1 ns. At 100 K, a rigid active site structure of ChR-2 is obtained. However, in two out of 14 trajectories transitions between two structural patterns have been observed. The structural patterns are summarized in the appendix in Table 1.

Comparing the starting structure at the beginning (300 K) with the structure at the end 100 K, a higher amount of $RSBH^+...^-O$-E123 hydrogen bond motif in the trajectories has been found. These significant structural changes happened during the cooling process. Thus, the $RSBH^+...^-O$-E123 is favored at low temperatures and from a statistical point of view, this indicates a slightly lower free energy of the $RSBH^+...^-O$-E123 motif (ca. 2 kcal/mol).

Furthermore, in the solid-state NMR-study[50] a smaller distance between the $RSBH^+$ and the counter ion than in BR has been reported. The preference for the $RSBH^+...^-O$-E123 pattern results in the decrease of the distance between $RSBH^+$ and E123 carboxyl group. In Table 6.4.1 the averaged distances of the center-of-mass (of the carboxyl group) of the counter ions E123 and D253 to the $RSBH^+$ are listed. The averaged distance for the simulations at 300 K is only slightly reduced compared to the distance measured in the BR trajectory. Upon closer analysis, the distances of the low temperature trajectories a larger decrease is observed. This mainly results from the preferred $RSBH^+...^-O$-E123 and the less heterogeneous hydrogen bonding pattern of the $RSBH^+$.

The simulations at low temperature sampled a decrease of the HC14-C15H dihedral (167°) of about 8° compared to the simulations at 300K (175°) (see Table 6.4.2). This temperature dependent alteration of the dihedral angle in the simulations gives the hint

Table 6.4.1: The average distance between the $RSBH^+$ nitrogen atom and the center of mass of E123 side chain carboxyl oxygens for the ten trajectories sampling the $RSBH^+ ... {}^-O$-(E123) pattern [Å] (standard deviation).

	BR-xray (1R2N)	BR-trajectory	C1C2-Xray (3UG9)	ChR-2 all-*trans* at 300 K
$RSBH^+$-^-OOC(E123)	3.66 (OD2) 5.29 (OD1)	3.83 (0.38) (COM)	3.44 (OE1) 3.94 (OE2)	3.66 (COM)
$RSBH^+$-^-OOC(D253)	3.30 (OD1) 4.79 (OD2)	4.69 (0.44) (COM)	3.01 (OD2) 3.66 (OD1)	4.65(COM)

cooling 100K	1	2	3	4	5
$RSBH^+$-^-OOC(E123)	2.76 (0.09)	2.73 (0.09)	2.79 (0.07)	3.24 (0.09)	3.33 (0.09)
$RSBH^+$-^-OOC(D253)	4.37 (0.21)	4.57 (0.22)	5.04 (0.16)	5.36 (0.13)	5.15 (0.16)
cooling 100K	6	7	8	9	10
$RSBH^+$-^-OOC(E123)	3.74 (0.18)	2.99 (0.11)	2.83 (0.07)	3.07 (0.13)	2.75 (0.07)
$RSBH^+$-^-OOC(D253)	5.31 (0.15)	5.20 (0.16)	4.00 (0.19)	5.71 (0.14)	4.41 (0.18)

that the experimental value of the solid-state NMR measurements (158°)[50] at 80K does not accord to the value at 300K. The dihedral angle value at 100K is almost the same as the experimentally measured dihedral angle in BR (164°)[126] and Green Proteorhodopsin (GPR) (161°)[127] (see Table 6.4.1). The good reproduction of the low temperature HC14-C15H dihedral is a hint that the current QM/MM sampling at DFTB3 level gives a representative retinal geometry, suggesting an almost planar all-*trans* retinal and a ca. 20° twisted 13-*cis* retinal around the C14-C15 bond at 300 K for ChR2-WT and ChR2-C128T complexes. Moreover, comparison of the stability of the hydrogen bond at the DC-Gate shows, that the water-bridged hydrogen bond is more stable at 100K than at 300K. The percentage of water bridged DC-Gate increases (47%-90%) in simulations at 100 K in comparison to simulations at 300K (30-55%).

Table 6.4.2: Averaged CC13=C14C, HC14-C15H, CC14-C15C torsional angles

	Retinal isomer	Torsional angle at 300K (100K) [°]	
		CC13=C14C	HC14-C15H
ChR-2 (water-bridged DC gate)	all-trans	178 (175)	175 (167)
	13-cis	12	197
ChR-2 (direct DC gate)	all-trans	184	182
	13-cis	-	-
ChR2-C128T	all-trans	179	179
	13-cis	15	200
BR	all-trans	189	184
Isolated retinal	all-trans	180	180
	13-cis	0	180
BR-xray (1C3W)[128]	all-trans	203	-
BR-xray (1X0S)[129]	13-cis	2	-
C1C2-xray (3UG9)[30]	all-trans	180	-
GPR-xray (4JQ6)[130]	all-trans	180	-
ChR2 exp.[50]	all-trans	-	158
BR exp.[126]	all-trans	-	164
GPR exp.[127]	all-trans	-	161

6.4.5 Absorption Spectrum

Comparing the experimental spectra of BR and ChR-2, they differ in wavelength and in the number of peaks. While the BR spectrum shows only one peak, in ChR-2 the spectrum is blue shifted about 0.45 eV and three main peaks are present, at 2.62 eV, 2.80 eV and 2.99 eV (see Figure 6.1.1). The ChR-2 spectrum is recorded using a dark-adapted ChR-2 (after several minutes in darkness) with 70:30 mixture of all-*trans* and 13-*cis*, 15-*syn* retinal.

Initially, the blue shift of the ChR-2 spectrum has been suggested to result from the strong

hydrogen bonding interaction between the $RSBH^+$ and the ionic counter ions E123 and D253, the so-called salt bridged motif. It was hypothesized, that it could stabilize the S0 state of the retinal, leading to a blue shift of the excitation energy compared to BR.

However, another study concluded, that the main shift derives from the different amino acid composition in the binding pocket of ChR-2 and BR and, by extension, the changed electrostatic interaction between the binding pocket and the retinal. The hydrogen bonding interaction at the $RSBH^+$ is only important for the absolute shift of the absorption property of retinal, but does not have so much influence on the shift between BR and ChR-2.[98,99]

The explanation for the single peak character in the BR absorption spectrum is a rigid active site structure (pentagonal cluster) representing a global ground state minimum with harmonic oscillations around this minimum, due to thermal effects, leading to homogeneous spectral broadening.[119]

However, as described in the previous section, ChR-2 has not such a global minimum structure, in fact the active site can adapt one of several ground-state conformations, which could explain the inhomogeneous spectral broadening of the absorption spectrum.

SORCI calculations on 50 QM/MM-minimized randomly chosen snapshots of the trajectories, in a former study by the author,[115] confirmed the assumption that the heterogeneity of the structures is responsible for the broad excitation energy, since the snapshots did not collapse in one global minimum after minimization. Furthermore, the SORCI spectrum showed a range from 2.85 eV to 3.00 eV. This matches very well with the experimental spectrum of ChR-2. Note, that in BR, the QM/MM minimization and SORCI calculations lead to only one structural pattern, the pentagonal cluster, and only one excitation energy. Nevertheless, the spectral shift of 0.45 eV between BR and ChR-2 is reproduced by the SORCI calculations (about 0.50 eV[115,121]), while the absolute value of the excitation energy cannot be reached due to the lack of polarization effects of the surrounding protein.

Due to the structural variety of the ChR-2 active site a simpler method should be used to describe dynamical effects on the absorption spectrum. The semi-empirical method OM2/MRCI turned out to be suitable for the retinal excitation after an extensive benchmark by Beatrix Bold and Yanan Guo.[115,121] The OM2/MRCI method has been used to calculate the excitation energy of 13 000 snapshots (7 000 of all-*trans* and 6 000 of 13-*cis*, 15-*syn* retinal) along the trajectories. In the spectrum in Figure 6.4.6 the excitation energies are summarized in a histogram. Fitting a Gaussian function to the data ($g(x) = a_0 e^{a_2(x-a_1)^2}$) gives the maxima of the spectrum. Furthermore, to allow comparison the histogram of 1000 snapshots from the BR trajectory has been computed as well. The fitted maxima show a red shift of about 0.57 eV compared to ChR-2 (see Figure 6.4.6). This is slightly higher than the experimental spectral shift, but within the accuracy of the OM2/MRCI method.

Regarding the calculated spectra of all-*trans* and 13-*cis*, 15-*syn* retinal in ChR-2 (see

Figure 6.4.6) the different retinal conformations lead to a spectral shift of about 0.05 eV, which is only a minor shift compared to the all-*trans* retinal in the absorption spectrum. Since the 13-*cis*, 15-*syn* trajectories show similar structural patterns as the all-*trans*, the inhomogeneous absorption spectrum could result from both retinal configurations. However, the question if the two retinal configurations are both present in ChR-2 can not be solved in this study.

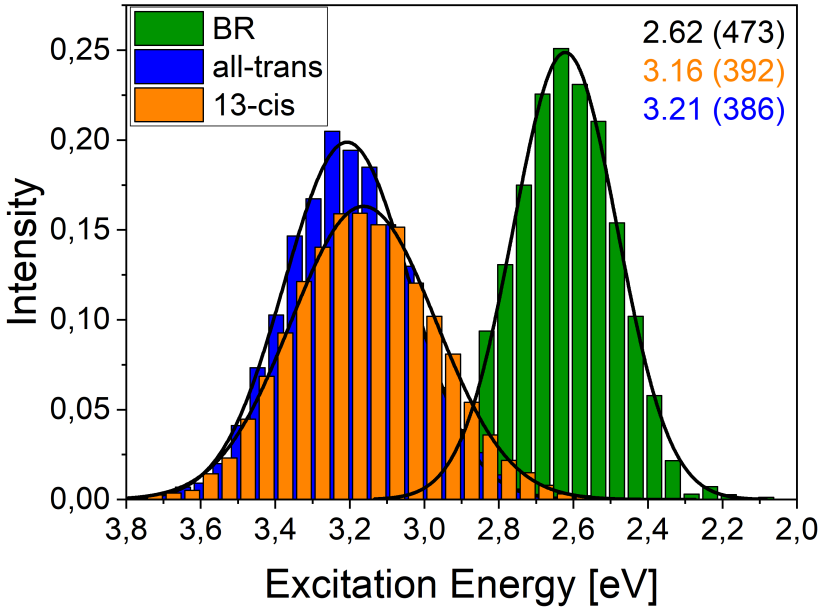


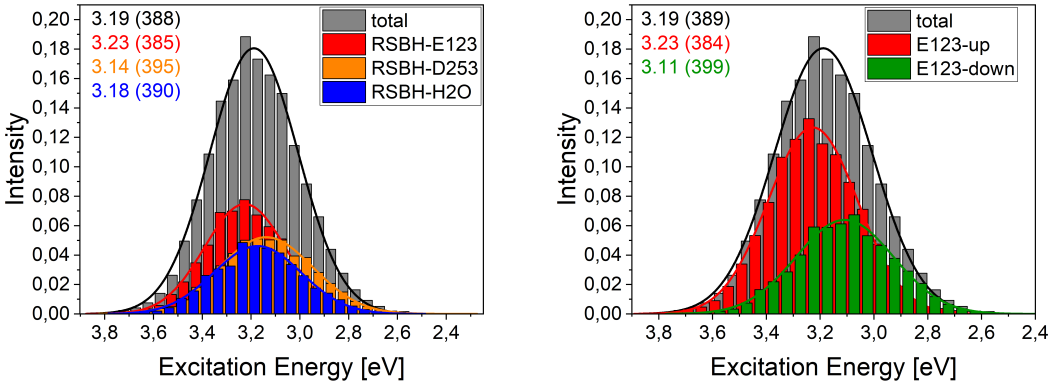
Figure 6.4.6: Calculated absorption spectrum of ChR-2 bound to all-*trans* retinal (blue) and 13-*cis*, 15-*syn* retinal (orange) and BR (green). The absorption maxima in eV are computed by fitting a Gaussian function. In parentesis are shown the maxima in nm

Analyzing the influence of the active site structural patterns on the absorption spectrum, the histograms have been divided on the basis of the particular structural $RSBH^+$ motif (see figs. 6.4.7a, 6.4.7c and 6.4.7e) and in the two E213 side chain conformations (see figs. 6.4.7b, 6.4.7d and 6.4.7f). The maximum shift between the spectrum of the $RSBH^+$ hydrogen bonding patterns is about 0.09 eV between $RSBH^+$...$^-$O-E123 and $RSBH^+$...$^-$O-D253 motif for the ChR-2 bound to a mixture of all-*trans* and 13-*cis*, 15-*syn* retinal. Considering only the all-trans retinal the maximum shift is between the $RSBH^+$...$^-$O-E123 and the $RSBH^+$..._O-H2O patterns and reduced to 0.05 eV, whereas in the protein bound to 13-*cis*, 15-*syn* retinal the maximum shift is between $RSBH^+$...$^-$O-E123 and $RSBH^+$...$^-$O-D253 patterns of about 0.09 eV.

The E123 side chain conformation shifts the spectrum about 0.12 eV for the protein bound to the mixture of retinal, 0.07 eV for the protein bound to pure all-*trans* and 0.15 eV for the protein bound to pure 13-*cis*, 15-*syn* retinal. The shift in the experimental absorption spectrum of ChR-2 is about 0.18 eV and 0.37 eV between the main and sub peaks. Therefore, it can be concluded, that within the accuracy of the method, both retinal

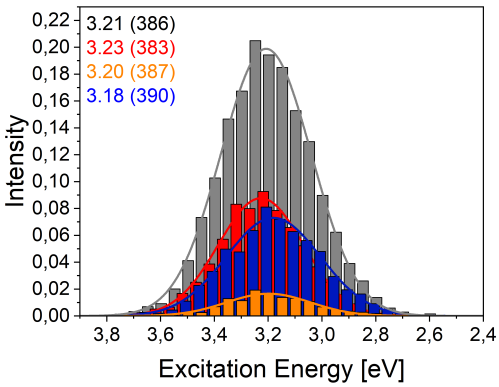
conformations (all-*trans* and 13-*cis*, 15-*syn*) as well as the structural variety of the E123 side chain conformation contribute to the fine structure of the absorption spectrum in ChR-2.

In particular the E123 side chain conformation is the key structural reason for the fine structure, since this pattern leads to the largest spectroscopic shift (0.15 eV).

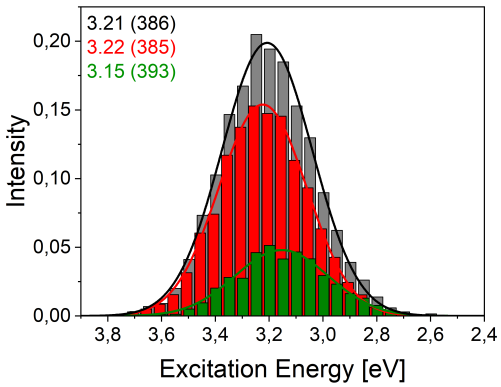


(a) all-*trans* & 13-*cis*, 15-*syn* retinal

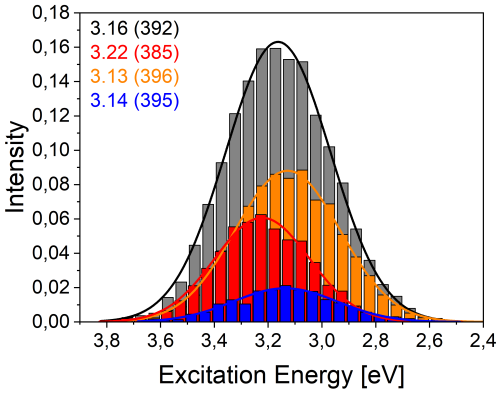
(b) all-*trans* & 13-*cis*, 15-*syn* retinal



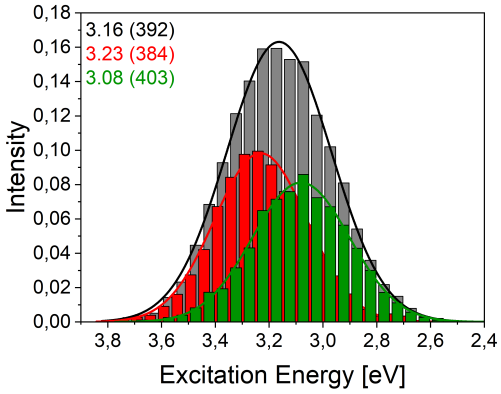
(c) all-*trans* retinal



(d) all-*trans*retinal



(e) 13-*cis*, 15-*syn* retinal



(f) 13-*cis*, 15-*syn* retinal

Figure 6.4.7: Calculated absorption spectrum of trajectories in the form of a histogram containing all-*trans* & 13-*cis*, 15-*syn* (single point excitation energy calculations with OM2/MRCI of 13 000 snapshots) Contribution of the different hydrogen bond patterns (a, c, e) and contribution of the different E123 side chain conformations to the spectrum (b, d, f) in trajectories containing a mixture of both retinal conformations (a, b), all-*trans* (c,d) or 13-*cis*, 15-*syn* (e, f). Gaussian functions are used to fit the spectra. The corresponding maxima are listed in the same color in eV (nm). In total all 13 000 snapshots are concentrated in the histogram.

6.4.6 DC-Gate

In the ChR2-C128T mutant the C128 is replaced by a threonine amino acid. This mutation leads not only to a 200- to 1000- fold extended lifetime of the conducting state,[56] but also to a direct hydrogen bond between the T128 and the D156 residues.[102,121] Yanan Guo performed a detailed computational study on the ChR2-C128T mutant[121] as described above for the ChR-2 wild type (WT).

The hydrogen bond pattern in simulations with a water bridged DC gate (containing 13-*cis*, 15-*syn* and all-*trans* retinal) corresponds to 40 % RSBH-E123, 33 % RSBH-D253 and 27 % RSBH-H2O, while in the simulations of ChR2-C128T with direct DT gate the hydrogen bond pattern is composed of 72.1 % RSBH-E123, 4.5 % RSBH-D253 and 23.4 % RSBH-H2O (containing 13-*cis*, 15-*syn* and all-*trans* retinal). Based on these results, the question has arisen, whether there is a coupling between the structural pattern of the DC-Gate and the hydrogen bond pattern of the active site. Does the percentage of the hydrogen bond pattern change with respect to the DC-Gate pattern?

As mentioned above in section 6.3, two different DC-gate structures have been found in the MM equilibration simulations of the ChR-2 models. On the one hand a water molecule bridges D156 and C128 (water bridged DC-gate, Figure 6.4.8a) and on the other hand a direct hydrogen bond is formed between D156 and C128 (direct DC-gate, Figure 6.4.8b), where D156 serves as a donor and C128 as an acceptor. Up to here the direct DC-Gate structure has been discarded from the analysis, since the vibrational frequency calculations indicate only a water-bridged DC-Gate in ChR-2 (see section 6.3). Furthermore, the water bridged DC gate has been proposed by previous classical MD simulations[42] and validated by QM/MM vibrational frequency calculations.[102] Controversially, spectroscopic measurements suggest a direct bond DC-Gate.[39,49,64,101] The hydrogen bonding situation in the DC-Gate was not solved for the ChR-2 Dark-state and was a controversial issue. In the discarded QM/MM simulations with a direct DC-Gate, the active site structures pattern composition differs compared to the water bridged simulations, thus a coupling between the DC-Gate motif and the active site has been assumed. To further analyze the active site of the structure with a direct DC-Gate and its subsequent impact on the excitation energy, the discarded direct DC-Gate structures are reactivated and the same QM/MM simulations as for the water bridged DC-Gate have been performed and analyzed.

The behavior of the direct-DT-Gate structure in ChR2-C128T of Yanan Guo's study and the water bridged DC-Gate structure can be directly compared with the obtained results on the direct DC-Gate in the wild type. Furthermore, the correlation between the water bridged DC-Gate, the direct DC-Gate and the direct DT gate with the active site is analyzed.

Six QM/MM MD simulations on ChR-2 containing a direct DC-Gate have been performed with a timescale of 1ns. As starting structure for the QM/MM simulations, the discarded

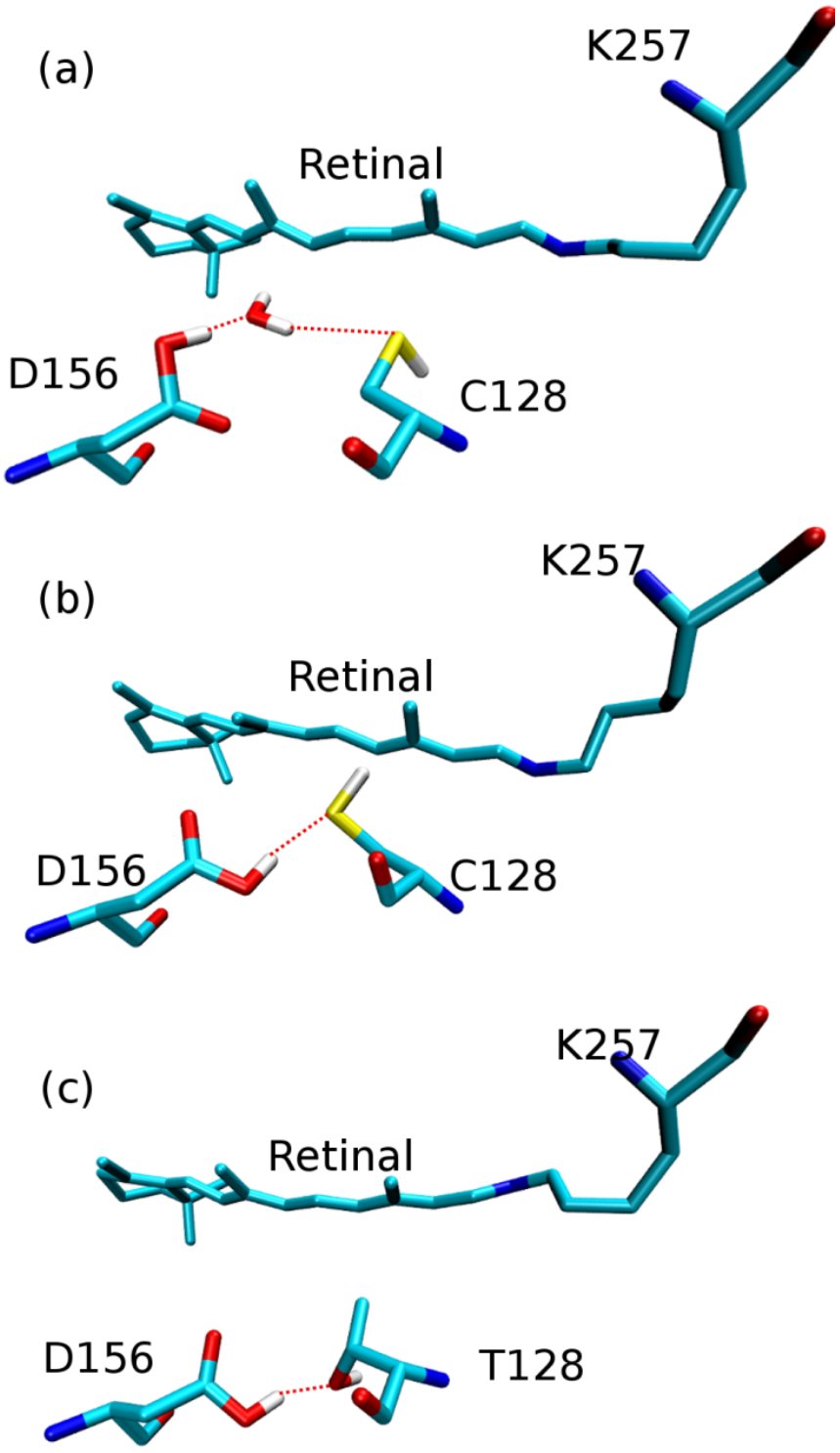


Figure 6.4.8: Different DC-Gate structures after Equilibration. a) water bridged DC-Gate and b) direct DC-Gate of ChR-2 and c) the direct DT-Gate in ChR2-C128T mutant adapted from.[121] In the following productive QM/MM simulations the water bridged DC-Gate and the DT-Gate are stable, while the direct DC-Gate forms different motifs during the simulations (see text and Figure 6.4.9)

structures after the equilibration step (see section 6.3), containing a direct DC-Gate have been used. For the analysis of the water-bridged DC-Gate the previously described trajectories are used and for the analysis of the DT-Gate the trajectories by Yanan Guo have been used[121]

The analysis of the DC-Gate hydrogen bond pattern of the QM/MM simulations containing the water bridged DC-Gate show a stability of the water bridged hydrogen bond between C128 and D156 about 32,6%. However, the carbonyl O-H of D156 is hydrogen bonded to the oxygen atom of a nearby water and this hydrogen bond is stable about 98,8 % in the trajectory. Neither a turn of D156, nor a change in the dihedral of the C128 side chain within 5 ns simulation time is visible. This hydrogen bond pattern is stable in all QM/MM simulations. The S-H group of C128 points mainly to T127, but no hydrogen bond between C128-S-H and O-T127 (Figure 6.4.8a) is formed. However, the S-H group is very flexible and rotates several times during the simulation, this is in accordance to vibrational frequency studies.[131]

In the simulation of the ChR-2-C128T mutant with the DT-Gate a hydrogen bond of O-H T128 to the backbone carbonyl group of W124 residue is observed as well as a turned carboxyl group of D156 with respect to the orientation in the water bridged DC-Gate structure (see Figure 6.4.8). This structure motif is stable in all six simulations.

In the simulations containing a direct DC-Gate, three different types of structure motifs have been observed:

i) In the first motif (motif 1), the carboxyl group of D156 is turned with respect to the water bridged-DC-Gate. A hydrogen bond between the carboxyl (C=O) group of the D156 side chain and a nearby water molecule is formed and the S-H group of the C128 side chain points in the direction of the retinal (see 6.4.9a). This DC-Gate motif has been observed most frequently, in four out of six simulations.

ii) In the second motif (motif 2 in 6.4.9b), the carboxyl group of D156 side chain orients like in the water bridged DC-Gate motif, the hydroxyl group of D156 side chain points next to retinal chromophore, and the S-H group of the C128 side chain points towards the retinal.

iii) In the last DC-Gate motif (motif 3 in 6.4.9c) the hydrogen bond pattern is like in the DT-Gate motif. The carboxyl group of D156 side chain is turned with respect to motif 2, but the S-H group of C128 side chain forms a hydrogen bond to the backbone carbonyl group of the W124 residue. In simulations containing motif 3, the E123 side chain rearranges in the downward orientation, this is not the case in the trajectories containing motif 1 or 2. Motif 2 and motif 3 have been observed respectively only in one of six trajectories.

Since six independent simulations have been performed, each structure motif is stable at least over 1 ns. The S-H group of the C128 side chain does not form any hydrogen bond,

only in motif 3 a hydrogen bonded S-H group is observed. And only in simulations with motif 3 a 15% to 30% higher probability for a E123 downward conformation can be observed. Therefore, the hydrogen bond of the S-H group of the C128 side chain seems to influence the E123 conformation. This is supported by the fact, that in the ChR2-C128T mutant, the E123 downward orientation is also sampled more often compared to the ChR-2 WT containing water bridged DC-Gate and hydrogen bond pattern in ChR2-C128T mutant is similar to the motif 3 in the direct DC-Gate.

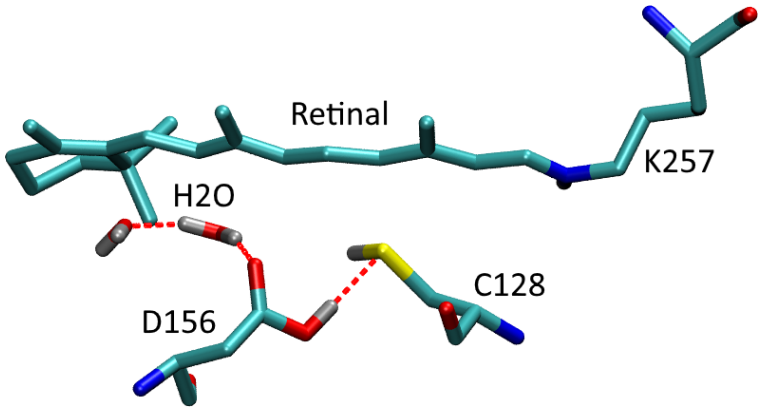
The simulations with the water-bridged DC-Gate sample a relatively rigid structure, no changes in the hydrogen bonding pattern can be observed, thus the pattern is stable in all simulations, while in the direct DC-Gate three different hydrogen bond motifs are observed. The variety of the hydrogen bonding patterns in the direct DC-Gate simulations could indicate a structural origin for the fine structure in the absorption spectrum or it is an indication for the reduced stability of this DC-Gate structure.

QM/MM Simulations with DC-Gate in the QM-zone

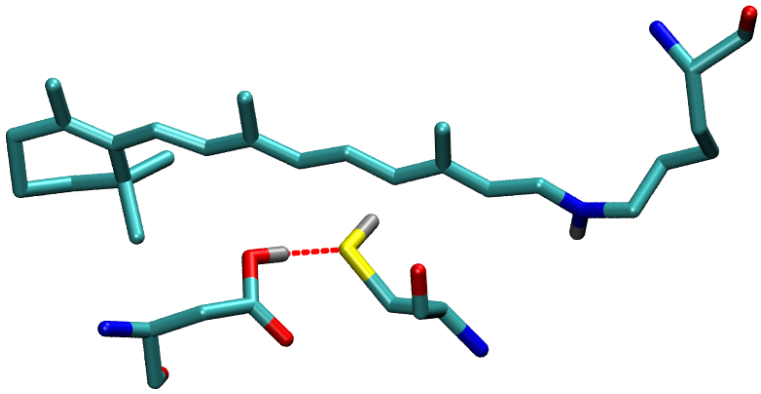
Since with classical force fields mainly a salt bridge between the retinal SB and the counter ion E123 can be sampled, so a detailed description of the chromophore geometry can be achieved only by quantum mechanics, therefore QM/MM simulations have been performed. Indeed, it was shown in the former sections, that the active site of ChR-2 can be properly described only by including the retinal and its counter ions in the QM region. But what about the hydrogen bond pattern at the DC-Gate? Will the hydrogen bonds be described correctly, if D156 and C128 are not included in the QM-zone?

For further testing of the hydrogen bond pattern at the DC-Gate, nine QM/MM simulations containing the DC-Gate, the water bridged DC-Gate and in case of the water bridged DC-Gate with an additional water molecule in the QM-zone have been performed, each with a simulation time of 1 ns. As starting structure for four trajectories the water-bridged DC-Gate motif and for five trajectories the direct DC-Gate has been chosen. The starting motif of the direct DC-Gate is either motif 1 or motif 2 as shown in figs. 6.4.9a and 6.4.9b.

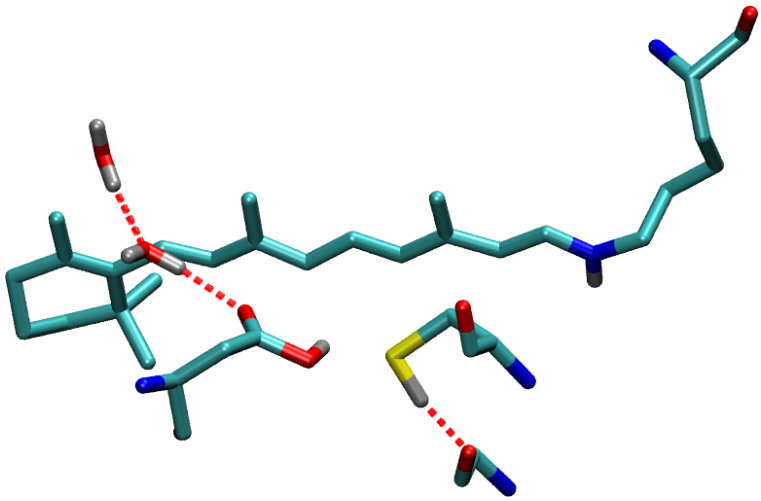
The trajectories with the direct DC-Gate show, that in two out of five trajectories, a water molecule moves in the DC-Gate within 1 ns simulation time and stays there for the rest of the simulation. Similar reorientations of the D156 and C128 side chains happen before a water molecule moves between D156 and C128. The D156 carboxyl group switches, so that the carbonyl group of D156 is now next to the retinal molecule, then the C128 side chain reorients and adopts the conformation pointing to T127. The distance between the alcohol group of T127 and H-S group of C128 decreases from 6.5 Å to 4.5 Å, while the distance between D156 and C128 increases. In the last step, a water molecule moves in the gap between D156 and C128 and the same conformation as in the water bridged motif (Figure 6.4.8) is adopted, due to this, the distance between S-H of C128 and side chain O-H



(a) Direct DC-Gate motif 1



(b) Direct DC-Gate motif 2



(c) Direct DC-Gate motif 3

Figure 6.4.9: Direct DC-Gate motifs found in six QM/MM simulations. Motif 1 has been observed in 3 out of 5 simulations, while motif 2 and motif 3 are found only in one simulation.

of T127 decreases again. In the simulations where no water molecule moves in, mainly the motif 2 shown in Figure 6.4.9b is sampled. This indicates that motif 2 is the most stable hydrogen bonding pattern motif and motif 1 and motif 3 arise in the QM/MM simulations without the DC-Gate in the QM-zone only due to the incorrect description of hydrogen bonds by the force field. Thus, further analysis of the trajectories with a direct DC-Gate is based on the simulations containing the DC-Gate in the QM-zone.

In the four trajectories starting from the water bridged DC-Gate, the hydrogen bond pattern of the DC-Gate stays stable. There are no structural changes over 1 ns simulation time. The average distance between S-C128 and O-T127 is about 4.3 Å and the dihedral angle of C128 side chain is about 145°. Therefore, the results from the former analysis can be used for the comparison with the trajectories containing a direct DC-Gate.

Correlation Between the DC/DT-Gate and the Active Site

A different active site hydrogen bond pattern in the direct DC-Gate simulations is observed compared to the water bridged DC-Gate. The percentage of the hydrogen bond pattern of the simulations with different DC-Gate is shown in Table 6.4.3. Here only the all-*trans* trajectories are considered. The hydrogen bond pattern $RSBH^+...OH_2$ is preferred in the direct DC-Gate while in simulations with a water bridged DC-Gate the pattern $RSBH^+...^-$O-E123 and $RSBH^+...OH_2$ are almost equally populated. The comparison of the ChR2-C128T simulations shows a similar distribution compared to the ChR-2 WT simulations with direct DC-Gate. This indicates a coupling between the DC-Gate and the hydrogen bond pattern at the active site. Nevertheless, the hydrogen bond pattern $RSBH^+..._$O-D123 is slightly more stable in ChR-2 WT than in the ChR2-C128T mutant. As suggested for the ChR2-C128T mutant,[115,121] the downward configuration of the E123 side chain leads to a main hydrogen bond pattern with $RSBH^+...OH_2$.

Thus, the movement of the E123 side chain leads to a structural difference in the active site pattern. In order to get a better insight in the correlation between DC-Gate and the active site, the E123 side chain conformations has been analyzed in the direct DC-Gate simulations (Table 6.4.3).

The direct DC-Gate of ChR2-C128T shows smaller population in the E123 upward orientation, thus it can be expected, that the E123 upward conformation is also less present in the direct DC-Gate of ChR-2. But the opposite has been observed in the simulations. The E123 upward conformation is preferred in direct DC-Gate of ChR-2 as well as in the water-bridged DC-Gate of ChR-2. This means the E123 upward conformation in direct DC-Gate in ChR-2 is more stable than in the DT-Gate of ChR2-C128T mutant. Therefore, the water in the DC-Gate is not the origin for the stability of the hydrogen bond between E123 side chain and T127 side chain. The orientation of the E123 side chain is not affected by the DC-Gate hydrogen bond pattern. Moreover, the different stability of the E123 side chain and

T127 side chain hydrogen bond could result from changed electrostatic interactions, when replacing the C128 with a threonine.

Table 6.4.3: Summary of structural properties of the active site with different DC-Gate hydrogen bond pattern. Experiment mutant: 2.58 eV Experiment ChR-2 wild type 2.62 eV.

		Water bridged DC-Gate	**DT-Gate**	**Direct DC-Gate motif 2**
Distance	D156-O-H – S-C128/T128	5.3 Å	2.8 Å	4.0 Å
	C(T)128-S-H – O-T127	4.3 Å	5.7 Å	5.2 Å
	T127-O - RSB	3.6 Å	3.6 Å	3.1 Å
Dihedral	C128 side chain	149.6°	–78.8°	92.7°
E123 side chain	up	76%	18%	81%
conformation	down	24%	82%	19%
RSBH hydrogen	-RSBH+. . . -O-(E123)	49%	19%	10%
bond pattern	-RSBH+. . . -O-(D253)	9%	9%	32%
	-RSBH+. . . OH2	42%	73%	58%
Excitation energy	OM2/MRCI	3.19 eV	3.08 eV	3.06 eV
Shift to BR	OM2/MRCI	0.57 eV	0.46 eV	0.44 eV

The results show a coupling between the DC-Gate structure and the hydrogen bond pattern next to the $RSBH^+$. A different hydrogen bonding pattern distribution has been sampled for the water-bridged and for the direct DC-Gate in ChR-2. To get a better insight in this coupling, the stability of the hydrogen bond between E123 and T127 side chain, which directly influences the E123 side chain conformation in the active site, has been analyzed. Thus, one can conclude that there is no direct coupling between the E123 side chain orientation and the DC-Gate motif. The coupling between DC-Gate and active site must originate from the electrostatic interaction and not from the hydrogen bond between E123 and T127 side chain.

Nack *et al.*[43] proposed a direct DC-gate (in ChR-2) or DT-gate (in ChR2-C128T mutant) structure, i.e. a direct hydrogen bond between the side chains of D156 and C128/T128. However, the simulations on the direct DC-Gate performed in this study suggest, that D156 serves as the donor (see Figure 6.4.9). Two observations are in favor of the D156 side chain acting as donor: (i) in the C1C2 crystal structure, the thiol group of C167 (C128 in ChR2) points away from D195 (D156 in ChR2);[42] (ii) the extended photocycle kinetics observed in ChR2 after C128T mutation can hardly be explained by the same DC and DT gate structure.

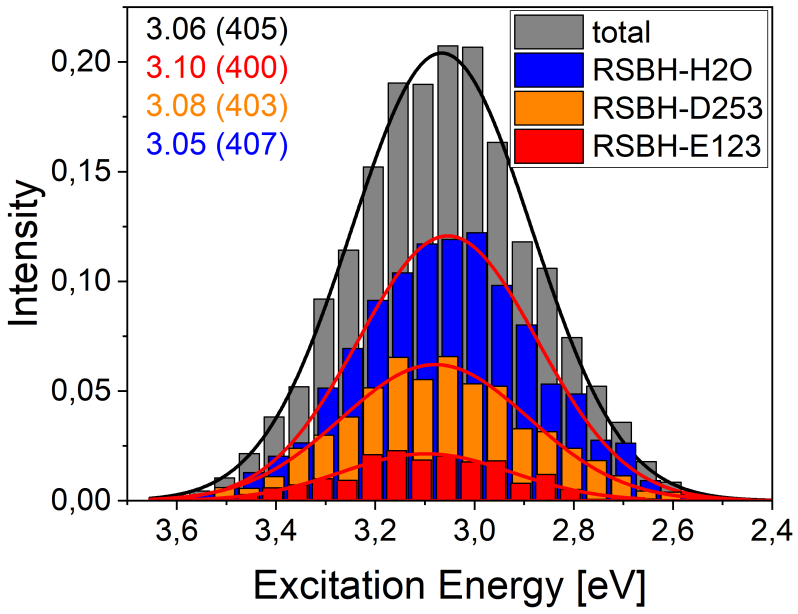
The D156 serves as proton donor for the RSB in the later states in the photocycle.[43,60] Therefore, the distances between the protonated carbonyl group of D156 and the $RSBH^+$ have been measured (see Table 6.4.3) in the simulations with different DC-Gate structure, to see whether a complete proton transfer from the D156 side chain to the RSB could be possible.

Excitation Energies

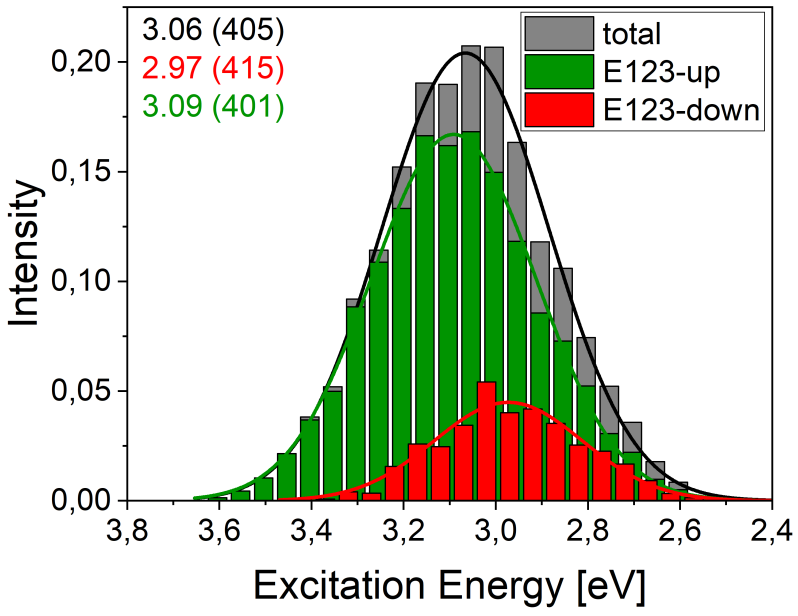
The slightly changed active site hydrogen bonding pattern found in the trajectories with a direct DC-Gate could lead to a shifted excitation energy of the retinal molecule. Thus, the excitation energies of 3 000 snapshots of the QM/MM trajectories (see Figure 6.4.11), containing the DC-Gate in the QM-zone and adopting motif 2, have been calculated with the OM2/MRCI method and have been analyzed in the same way as the water bridged DC-Gate in subsection 6.4.5.

The clustered spectra of the direct DC-Gate with motif 2 show similar excitation energy shifts between the structural patterns as in the water bridged DC-Gate trajectories (compare Figure 6.4.7).

The rearrangement in the surroundings of the retinal chromophore due to the water in the DC-Gate shifts the spectrum and yields a hypsochromic shift about 0.2 eV compared to the trajectories containing the water-bridged DC-Gate. Nevertheless, there is almost no shift between ChR-2 WT with direct DC-Gate and ChR2-C128T mutant. Thus, the change of the electrostatic environment due to the mutation has a minor influence on the chromophore conformation and the excitation wavelength.



(a) RSBH$^+$ hydrogen bonding pattern



(b) E123 side chain conformation

Figure 6.4.10: Calculated absorption spectrum of trajectories summarized in a histogram containing all-*trans* with the DC-Gate motif 2 in 6.4.9b and the DC-Gate in the QM-zone (single point excitation energy calculations with OM2/MRCI of 3 000 snapshots). a) Contribution of the different RSBH$^+$ hydrogen bond patterns and b) contribution of the different E123 side chain conformations to the absorption spectrum. Gaussian functions were used to fit the spectra. The corresponding maxima are listed in the same color in eV (nm).

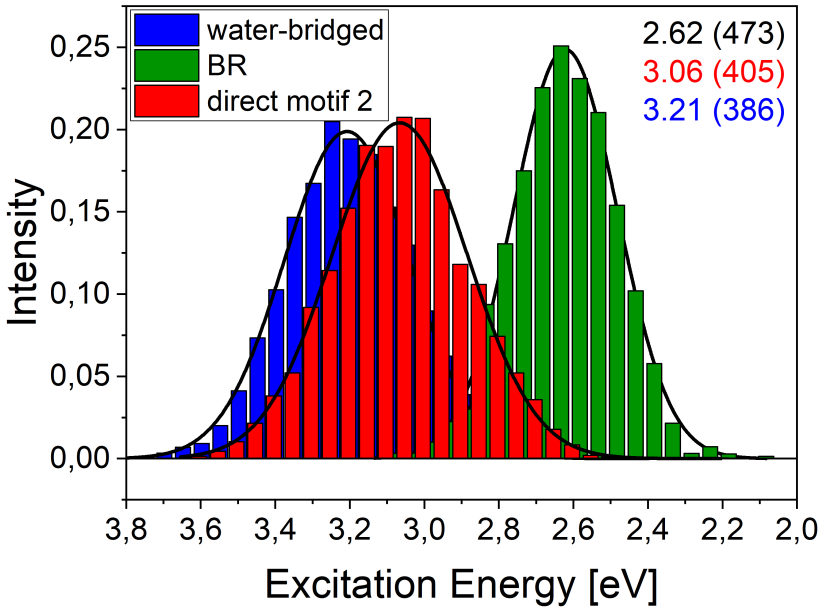


Figure 6.4.11: Calculated absorption spectrum of ChR-2 with water bridged DC-Gate (all-*trans*) (blue) and direct DC-Gate (red) and BR (green). The absorption maxima are computed by fitting a Gaussian function.

6.4.7 E90 Side Chain Configuration

Long MD simulations of the ChR-2 WT and ChR2-C128T mutant[121] with a timescale of 1 µs show, that the E90 side chain is flexible. After a simulation time of several ns (for details see appendix Table 2), the E90 side chain flips downward, towards the extracellular side, and the hydrogen bond partner is changed from N258 to D253 in all four trajectories of ChR-2. Interestingly, three out of four trajectories (each 400 ns) from the ChR2-C128T mutant show the flip of the E90 side chain, too.[121] In Figure 6.4.12 are shown three representative conformations of the E90 side chain, while the conformation in Figure 6.4.12a was only observed in the ChR2-C128T mutant and not in the ChR-2. These hydrogen bonding patterns are consistent with former studies on C1C2 chimera.[132] The backbone dihedral around C_β-C_γ of E90 is about $\pm 180°$ in the conformations shown in Figure 6.4.12a and Figure 6.4.12b, respectively and about 70° in the downward conformation (Figure 6.4.12c). In Figure 6.4.12c, the hydrogen bond is formed between E90 and the counter ion D253, while the N258 side chain amine forms a hydorgen bond with the S63 hydroxyl group.

The observed structural changes on the side chain dihedral angle of the E90 has been analyzed with well-tempered MM metadynamic simulation on the ChR-2 (see Figure 6.4.13), in order to get a free energy profile for this rotation. The profile shows a lower free energy for the downward orientation at 70° (Figure 6.4.12c) of the side chain dihedral angle of E90 than for the upward orientation at $\pm 180°$ (Figure 6.4.12b) and a barrier between the two conformations about 6.5 kcal/mol. That is to say, the downward orientation of the E90 side

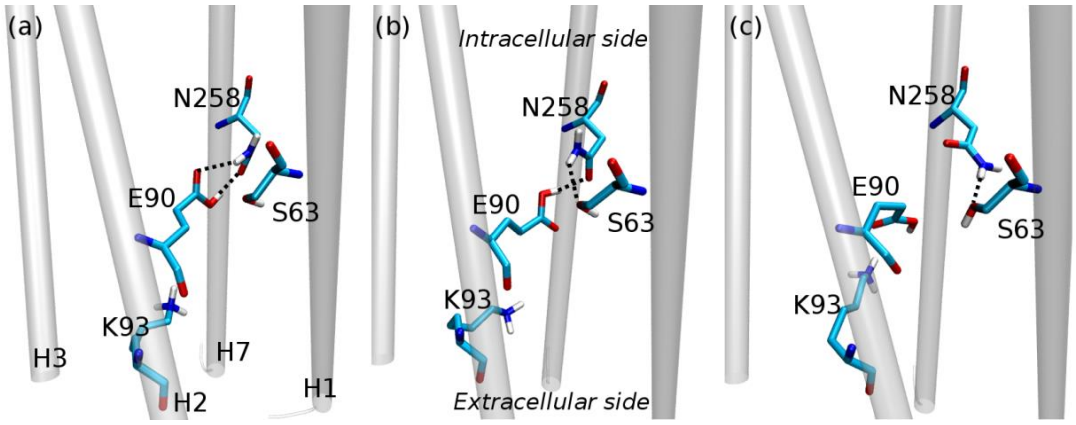


Figure 6.4.12: Three possible central gate structures. a) Two hydrogen bonds between E90 and N258; only observed in ChR2-C128T mutant; dihedral of E90 side chain of 180°; b) one hydrogen bond between E90 and N258 and one hydrogen bond between N258 and S63; observed in ChR-2 and ChR-2-C128T mutant; dihedral of E90 side chain of 180°; c) one hydrogen bond between N258 and S63; dihedral of E90 side chain of 70°. Figure adapted from.[121]

chain is preferred in the ChR-2 structure.

The timescale of the E90 flip in the 1 µs simulations is in range from 300 to 600 ns, this is consistent with the received barrier of 6.5 kcal/mol between $\pm$180° dihedral angle to 70° dihedral angle in the MM metadynamic simulation.

Therefore, the study of Kuhne *et al.*,[47] where a downward movement of the E90 side chain after 200 ns is suggested to be a first step of the photocycle after the isomerization, which leads to the channel opening, is an observed structure rearrangement due to the enlarged sampling on the homology model of ChR-2. Additionally, in Kuhne et.al[47] a strain is suggested to propagate from the K256 side chain to the N258 residue after the C13=C14 isomerization of the retinal. Therefore, it is suggested, that the hydorgen bond between N258 and E90 is weakened and this initiates the E90 downward movement after the isomerization. Interestingly, in this study the side chain of the retinal-linked Lysin (K256) is included in the QM region in the QM/MM simulations and it could be observed that the dihedral angle of the side chain is very flexible. This angle can change several times during 1 ns. Therefore, the propagation of the strain to the N258 side chain is not possible, because the strain can easily be converted to rotation energy of the K256 side chain.

Interestingly, a previous computational study[103] on a C1C2 homology model shows no downward rotation of the E90 side chain, however, the MM MD sampling time had been in the timescale of 4 µs. This could be caused by their use of the AMBER99-SB force field while in this study the CHARMM36 force field has been used for the MM MD simulations.

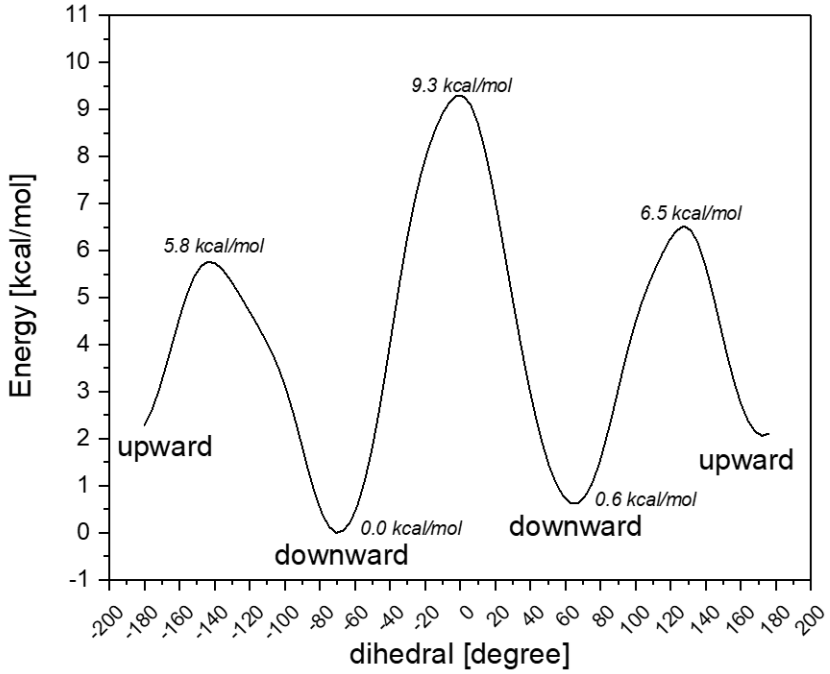


Figure 6.4.13: Free energy profile of E90 side chain rotation.

6.4.8 Simulations on Crystal Structure

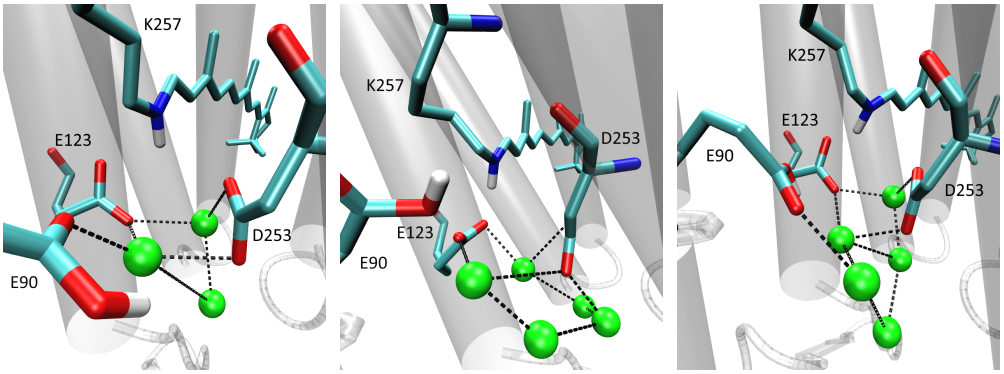
In late 2017 the crystal structure of ChR-2 was published (PDB-Code: 6EID[31]). The homology model used in this study is similar to the crystal structure. The major difference emerges in the orientation of the E90 side chain. As already discussed above, the E90 side chain is oriented upwards in the homology model and is thereby not in the energetically minimum (see Figure 6.4.13). This is also confirmed by the crystal structure, the E90 is downward oriented and points towards the D253 counter ion and builds up a water bridged hydrogen bond with the D253 and E123 counter ions.

Moreover, the crystal structure contains a water molecule between the side chain D156 and C128 and therefore confirms the water arrangements predicted both by this and several other computational studies.[42,103,115] Minor differences between the homology model used in this study and the crystal structure exist in the hydrogen bonding pattern between the two monomers of the dimer, as well as in the residues at the extracellular side of the protein (residues E101, E97 and Q56). Nevertheless, the effect of the E90 side chain orientation on the active site hydrogen bonding pattern and on the excitation energy needs to be analyzed. Therefore, 28 QM/MM simulations have been set up with four different QM-zones and six different water arrangements in the vicinity of the $RSBH^+$ (details see section 6.3).

Structural Characteristics

Concerning the water distribution in the active site, it should be noted that in the crystal structure three water molecules are present in the vicinity of the $RSBH^+$ (Figure 6.4.14a), while in the homology model structure, four or more water molecules are present (see Figure 6.4.14b).

The water molecules play an important role for the hydrogen bonding pattern at the $RSBH^+$. Since the crystal structure and the homology model differ in the water distribution inside the protein, the impact of the water amount on the hydrogen bond pattern needs to be analyzed. Therefore, QM/MM trajectories with the lower water density of the crystal structure and QM/MM trajectories with the higher water density (four or more water molecules) of the homology model have been performed and analyzed. In order to gain a starting structure of the crystal structure with a high water density, the water molecules of the homology model have been fitted and inserted into the crystal structure. Different water arrangements of several snapshots of the former trajectories of the homology models have been fitted to the crystal structure.



(a) (b) (c)

Figure 6.4.14: Water molecules (green) in the vicinity of $RSBH^+$ a) of the crystal structure b) of the homology model c) of MM MD of crystal structure after 600 ns simulation time.

Comparing the trajectories with the lower water density and the higher water density, it becomes clear, that the active site of the trajectories with crystal water molecules is much more rigid and fewer rearrangements happen within 1 ns simulation time. For example in the trajectories with the lower water density, the E123 side chain conformations stay in the upward orientation over the whole simulation time, while in the trajectories with the higher water density from the homology model, the E123 side chain adopts the downward orientation. However, the trajectories with the crystal structure water density are not as stable as the trajectories with the higher water density, since in several simulations, protonation of the counter ions E123 and D253 from the E90 side chain or nearby water molecules occur in the trajectories.

Furthermore, in the trajectories with higher water density, the downward orientation of the E90 side chain is not stable except in the trajectories with the largest QM-zone containing the retinal, K257, E123, D253, T127, K93, the DC-Gate and the E90. The E90 flips up within the first 10 ps in eight out of twelve trajectories, but if both the E90 and the DC-Gate are included in the QM calculation, the E90 is stable and stays in the downward orientation. However, in the trajectories with the smaller water density of the crystal structure the E90 downward orientation is stable in trajectories, no matter which QM-zone has been chosen. First and foremost, these findings imply the importance of the chosen QM-zone for the right description of such complex hydrogen bonding networks. Only the inclusion of the DC-Gate in the QM-zone stabilizes the E90 side chain at the downward orientation. Not even the trajectories with a QM-zone, containing the E90 without the DC-Gate prefer the E90 downward position, here in two out of four trajectories the E90 moves upwards. This is in contrast to the results from the MM metadynamics (see Figure 6.4.13), but a possible reason for this discrepancy could be the incorrectly described interaction between atoms in the QM region and the atoms in the MM region, since the electrostatic effects and polarization effects are considered only approximately (details see section 5.4) between these two regions. Since the trajectories with the largest QM-zone (retinal, K257, E123, D253, T127, K93, E90 and DC-Gate) include all effects between the relevant residues, the hydrogen bond pattern composition and retinal conformation in these trajectories as well as the distance between $RSBH^+$ and the counter ions have been analyzed (see Table 6.4.4).

Interestingly, the trajectory with the water distribution from the crystal structure (low water density), shows a salt bridge to the E123 side chain in only 12.6% of the snapshots, but the $RSBH^+$ does not form any hydrogen bond in most of the snapshots (86.2%). This is reproduced in the torsional angle of the retinal, which is almost planar. However, in the trajectories with higher water density, the structural pattern varies and the $RSBH^+$ forms hydrogen bonds with E123 (47.4%) side chain or a water molecule (22.1%). Furthermore, the E123 downward movement occurs in 37.0% of the snapshots and hardly ever in the low water density trajectories (0.3%). The averaged distance between $RSBH^+$ and any counter ion is shorter with a lower water density than with higher water density, but the torsional angle of the retinal molecule is with both water distributions almost planar.

With these results, it can be concluded, that the E123 salt bridged is not the only hydrogen bonding pattern formed by the $RSBH^+$, the structural variety analyzed in the former sections is therefore not wrong, but the diversity of the active site hydrogen bond pattern is strongly dependent on the water distribution at the active site. The water molecules are very important and play a key role for the structural arrangement in the active site.

The determination of the crystal structure takes place at very low temperatures around 80-100 K and the protein structures can adopt different structural patterns as been reported in previous studies.[113] More importantly, the water density can strongly decrease during

Table 6.4.4: Structural parameters over simulations with high and low water density in the crystal structure. The values for the high water density simulation are averages over five 1 ns QM/MM trajectories, while due to the instability of the simulations with low water density, the values are averaged over one 1 ns QM/MM simulation.

		high water density	low water density (x-ray)
excitation energy		3.23	3.23
E123 down flip [%]	down	37.0	0.3
	up	59.7	88.5
$RSBH^+$ Hbond pattern [%]	H2O	22.1	0.0
	E123	47.4	12.6
	D253	0.0	1.2
	no H-bond	30.5	86.2
torsional angle retinal [°]	CC13-C14C	178	182
	HC14-C15H	183	180
distance $RSBH^+$ to counter ions [Å]	E123-OE1	4.29	3.04
	E123-OE2	3.85	3.35
	D253-OD1	6.94	6.22
	D253-OD2	5.39	3.21

the crystallization procedure, since these are very harsh conditions for a protein, which are usually evolutionarily invested to perform optimally at a temperature range around 273-313 K (0-40 °C). This is the natural temperature range between frozen water and the denaturation temperature of proteins. Thus, due to the crystallization process at very low temperatures, unnatural rearrangements could happen and also the water density in crystal structures of proteins can be reduced.

To investigate the water amount at the active site at 300 K, a 600 ns long MM-MD simulation of the crystal-structure in a box with POPC-lipid bilayer and bulk water has been performed. The temperature of the simulation was set to 300 K. With the same conditions an additional MM-MD simulation of the crystal structure containing the water molecules of our homology model has been performed as well. The water distribution increases in the crystal structure after 600 ns. To illustrate this, the water distribution of the crystal structure (PDB Code: 6EID) is shown in 6.4.15a and the water distribution after 600 ns MM-MD simulation is shown in 6.4.15b. Furthermore, the water density at the vicinity of the $RSBH^+$ is increased as well (see 6.4.14c). The amount of water molecules at the active site after the long MM-MD simulation is comparable with the water molecule distribution in the homology model (cf. 6.4.14b and 6.4.14c). Therefore, the active site described in the former sections, with the high diversity on hydrogen bond pattern and side chain conformation is still reliable for ChR-2.

When starting a long MM MD simulation about 900 ns timescale with the crystal structure, containing the higher water density of the homology model, a slight decrease in the water density at the active site is observed (cf. 6.4.15c and 6.4.15d). Interestingly, the water density on the intracellular side of the protein (above the central gate) is reduced at the end of the simulation. It seems as if the the crystal structure pushes out the water on the extracellular side of the protein during the simulation.

Absorption Spectrum

The excitation energies over the trajectories have been calculated and there is no shift of the excitation energy (see Table 6.4.4) between the trajectories containing the crystal structure water distribution compared to the trajectories containing the higher water density distribution of the homology model. Comparing the averaged excitation energy of the simulations on the crystal structure (3.23 eV) with the excitation energy on the homology model (about 3.21 eV Figure 6.4.6), the influence of the E90 side chain orientation on the excitation energy is negligible and therefore does not shift the spectrum.

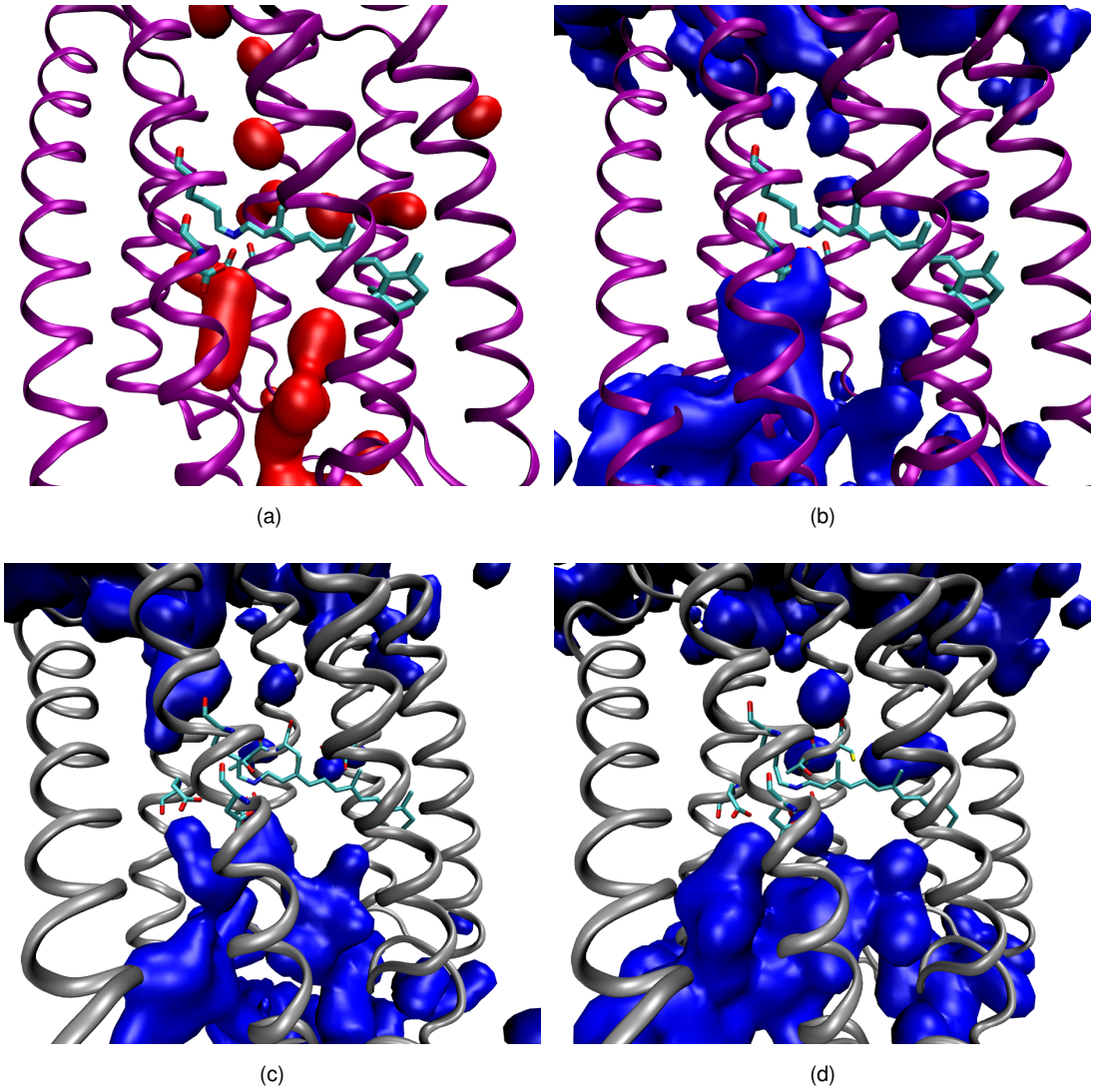


Figure 6.4.15: ChR-2 crystal structure with the water density distribution a) of the crystal structure (PDB code: 6EID[31]) (red) b) after a 600 ns MM MD simulation of the crystal structure (blue) and c) the water distribution of the homology model fitted in the crystal structure (blue) d) and the structure from c) after 900 ns MM MD simulation time.

6.4.9 Concluding Remarks

The detailed characterization of the binding pocket in ChR-2 is of great importance not only to rationalize color tuning studies, but also for the following reasons: i) the active site is crucial for the steps following the photoisomerization of the retinal, ii) the active site is directly connected to the central gate, which plays a key role in the ion selectivity of the channel and the pore formation during the photocycle iii) the DC-Gate structure pattern is important for the reprotonation of the RSB and the channel closing. All these points are crucial for the understanding of the function mechanism of the ChR-2. The detailed knowledge about the function of all the structure patterns enables a more directed and effective engineering of new optogenetic channels. For example the conversion of the ion selectivity to a K^+ or Ca^+ selectivity of the ChR-2.

In the study on the homology model a heterogeneous active site has been found, consisting out of three $RSBH^+$ hydrogen patterns and two E123 side chain conformations. Furthermore, a temperature dependency of the hydrogen bonding patterns has been uncovered. At 100 K the hydrogen bonding patterns are more rigid than at 300K, a preferred hydrogen bonding motif can be determined and therefore this motif is suggested to have a lower free energy than the other two patterns.

The absorption properties of this heterogeneous active site have been found to lead to the multi-peak nature of the absorption spectrum. Moreover, the hydrogen bonding patterns contribute to the broadening of absorption peaks and the E123 side chain shifts the excitation energy up to 0.1 eV and thus is responsible for the sub-peak nature of the spectrum. The impact of the 13-*cis*, 15-*syn* retinal conformation on the structural motif distribution at the active site has only been found in the statistical distribution of the patterns. Thus, the 13-*cis*, 15-*syn* retinal conformation does not change the absorption spectrum and therefore the 13-*cis*, 15-*syn* retinal conformation can not be excluded from the analysis due to computational reasons, but the question of its presence has been clarified experimentally now.[50]

A correlation between the presence of a water molecule in the DC-Gate and the active site hydrogen bonding pattern has been found. The inclusion of the DC-Gate in the QM-zone implied a water bridged DC-Gate structure, since in half of the QM/MM simulations the water molecule spontaneously moves in between D156 and C128 and the water bridged DC-Gate motif is stable over time, while in the other case several rearrangements happening during the simulations. It is astonishing, that within 1 ns simulation time in two out of five simulations a water molecule moves in the DC-Gate and stays there for the rest of the time. This is a strong indication for the presence of the water molecule in the DC-Gate. The simulations with different QM-zones show different results, this demonstrates the importance of a well-chosen QM-zone while describing the sensitive hydrogen bond patterns of proteins.

The direct DC-Gate shifts the absorption spectrum by about 0.15 eV compared to simulations with the water-bridged DC-Gate. This shift does not result from an E123 downward conformation, since the E123 upward conformation is more stable in the direct DC-Gate. Interestingly, the water in the DC-Gate shifts the excitation energy about the same value as the mutation of the C128 to T128. Thus, the electrostatic interaction in the ChR-2-C128T mutant due to the mutation in the vicinity of the retinal has to be canceled by other effects or has to have a minor impact on the excitation energy.

The controversially discussed E90 movement found by MM MD simulations is clarified by free energy calculations and it was determined that the E90 downward orientation is energetically preferred in ChR-2. Since the energy barrier between the upward and downward movement is in a range of several kcal/mol for ChR-2, long MD simulations with a time scale of several hundred ns or µs were necessary to detect this conformational detail.

In general, force field based methods are very useful tools to describe unknown protein structures and predict structural dependencies, but one has to be very careful with interpretations on homology models. An extended testing on structural arrangements is indispensable, not only for the reason that homology models are still approximated structures and can predict wrong conformations, but also if the models are very trustworthy, energy barriers of few kcal/mol between two structural states lead to rearrangements after an extended sampling time. This yields to the general and clear suggestion for extended sampling on homology protein models, since minor side chain rearrangements can be detected and with free energy calculations on these arrangements the model can be validated and further improved.

The computational methodology as described in this chapter has been frequently used in former studies on retinal proteins and the effect of different retinal conformations and protein structures can be well reproduced. Since the relative excitation energies between ChR-2 and BR are well reproduced, it can be concluded that the homology model represents ChR-2 structure.

Nevertheless, the free energy differences of the structural motifs found in this study are within a few kcal/mol, thus very small inaccuracies can shift the occupations of the states. However, the required accuracy is definitely higher than the chemical accuracy of 1 kcal/mol of the QM methods. But such high accuracy is difficult to achieve for standard electronic structure methods, that is why the picture shown here is only a qualitative picture and should not be over-interpreted.

Since in late 2017 the crystal structure was published, the suitability of the homology model used in this study could be tested. The crystal structure confirmed the results for the E90 side chain conformation calculated by free energy calculations.

The description of structural patterns within proteins, particularly hydrogen bonds, are

still a challenge for molecular force fields. The lack of polarization effects can lead to the underestimation of hydrogen bonds and to artifacts. QM methods perform better in describing hydrogen bonds and polarization effects, but their drawback is the computational cost, which brings up other challenges like size of the QM part and sufficient sampling of the protein. As was shown in the last section, the selection of the residues described by QM methods is crucial to get reliable results.

QM/MM simulations on the active site showed a correlation between the amount of water in the protein and the structural diversity in the binding pocket. With a higher water density at the active site, the active site hydrogen bond patterns are still very flexible as was found in the study on the homology model. Since the water distribution of the crystal structure of ChR-2 has been found to increase in 600 ns MM-MD simulations and at the active site more water is present than in the crystal structure, it leads to the conclusion that the high variety of structural patterns at the active site is still a reliable result.

7 | Modeling HKR1 Structure, Spectra and Function

Rhodopsins are light-sensitive receptor proteins, they respond to light and enable the signal pathways of the cells. The discovery of the Channelrhodopsins in the last decade paved the way for a new technology in the field of neuroscience. These light-gated ion channels enable neuroscientists to selectively activate nerve cells in tissues with short laser pulses. This technique is called Optogenetics and is a milestone for the investigation of neural networks. However, the inactivation of the nerve cells is not yet possible, especially by moderate light intensities. One solution might be provided by the recent discovery of the bimodal switching Histidin Kinase Rhodopsin 1 (HKR1) from the green alga *Chlamydomonas reinhardtii*. In detail, irradiating the protein with two different wavelengths makes it switch between two stable states. Bimodal switchable proteins have been found before, i.e. *Anabaena* Sensory Rhodopsin (ASR) or Sensory Rhodopsin II (SRII). But in none of these rhodopsins the second state is stable over several hours like in HKR1 and this makes it very interesting for protein engineering. However, the mechanism has to be understood and clarified first in order to adapt this unique feature on other rhodopsins. Thereby rhodopsins with new features can be developed for applications in the Optogenetic field. The subject of this chapter is about the investigation of the Rh-Bl state and all-*trans* to 13-*cis*, 15-*syn* isomerization of the retinal. Since there is no crystal structure available several structure models (homology models) have been built and simulated on the molecular and quantum mechanic level in order to gain an equilibrium state. The spectroscopic characteristics have been calculated based on the homology models and have been compared to experimental spectra for a validation of the binding pocket composition. This validation is an important step to ensure the reliability of the homology model. Calculations in the excited state on the obtained structures have been performed and gave an insight into the reactions taking place during the isomerization of the retinal in the excited state. Mainly, the 13*cis*, 15*syn* retinal structure is obtained, which makes a further investigation of the second stable state RH-UV in principle possible. Most of the work has been done during my stay in the lab of Igor Schapiro, especially the excited state calculations. He has an established expertise and experience in characterization of photochemical reaction path ways in rhodopsins by using

the quantum mechanical method CASSCF (Complete Active Space Self-Consitent Field) and molecular dynamics simulations in the excited state.

7.1 Model Building

The first step in investigating the HKR1 photocycle is to build a model of the dark-state structure of HKR1 the Rh-Bl state. Since there is no crystal structure available, a homology model has to be build. Therefore, the webpage Swiss-Model (https://swissmodel.expasy.org/)[133134135] has been used.

HKR1 is the first bimodal rhodopsin with two over several hours stable states, all other known switchable rhodopsins show a second stable states with a lifetime of only several minutes. Therefore, HKR1 is a unique rhodopsin and thus, the sequence homology to other known microbial rhodopsins is difficult and slightly sketchy. Furthermore, the HKR1 is a new found rhodopsin and there is almost no structural information about the composition of the active site, neither mutational studies had been made to identifying residues, which are crucial for the mechanism, nor much information from IR or Raman spectroscopy studies are available. IR spectroscopy studies identified the D239 residues as proton donor for the $RSBH^+$ proton.[68,69] These circumstances makes it hard to build a homology model and requires a validation of the homology models. Therefore, several independent homology models have been build based on different templates and are compared to each other, in order to find out a consistent composition of the binding pocket and to reduce the error rate. The best matching results of the Swiss-model sequence alignment have been used as templates for the homology model. Templates are the PDB structure of *Anabaena* Sensory Rhodopsin (ASR) (PDB-Code: 1XIO[136]), Halorhodopsin (PDB-Code: 3VVK[137]), Sensory Rhodopsin II (SRII) (PDB-Code: 2KSY[138]), Bacteriorhodopsin mutant M56A (PDB-Code: 1PXS[139]), the *Acetabularia* Rhodopsin I (PDB-Code: 5AWZ[140]) and the from the same organism stemming Channelrhodopsin (PDB-Code: 3UG9[30]). The homology models are named with respect to the PDB-Code of the respective template. And are only considered in the four letter PDB-Code in the following text.

All the selected templates show a sequence similarity of about 30% to HKR1 and all are microbial rhodopsins. Note, a 30% sequence similarity is the lowest similarity to get reasonable result for a homology model. In Figure 7.1.1 is shown the list with the best results of the sequence alignment analysis in Swissmodel and a summary of the used templates. Based on this six rhodopsins in the list, six independent homology models have been built using amino acids 27-305 of the sequence (Q6WRU3) found at the UniProt database (http://www.uniprot.org). The amino acids 27 to 305 are suggested to be the transmembrane domaine and therefore assemble the rhodopsin domaine of the HKR1.

The amino acids in the binding pocket shown in Figure 7.1.2 are located on the similar

Template	Seq Identity	Oligo-state	Found by	Method	Resolution	Seq Similarity	Coverage	Description
3ug9.1.A	14.41	homo-dimer	HHblits	X-ray	2.30Å	0.28	0.81	Archaeal-type opsin 1, Archaeal-type opsin 2
2ksy.1.A	16.74	monomer	HHblits	NMR	NA	0.28	0.80	Sensory rhodopsin II
5awz.1.A	16.13	monomer	HHblits	X-ray	1.57Å	0.29	0.79	Rhodopsin I
3abw.1.A	17.84	homo-trimer	HHblits	X-ray	1.90Å	0.30	0.77	Halorhodopsin
3vvk.1.B	17.84	homo-trimer	HHblits	X-ray	2.30Å	0.30	0.77	Halorhodopsin
1pxs.1.A	18.69	monomer	HHblits	X-ray	2.20Å	0.30	0.78	Bacteriorhodopsin

(a) Swiss model Results[133][134][135]

ASR (PDB:1XIO)	**Halorhodopsin (PDB:3VVK)**	**SRII (PDB:2KSY)**
Monomer	Trimer	Dimer
X-Ray (resolution: 2.0 Å)	X-Ray (resolution: 2.3 Å)	Solution NMR
1 Counter ion	1 Counter ion	2 Counter ion
Anabaena (Cyanobacterium)	*Natronomonas pharaonis* (Archae)	*Natronomonas pharaonis* (Archae)

BR mutant M56A (PDB:1pxs)	**Channelrhodopsin (PDB:3UG9)**	**Rhodopsin I (PDB:5AWZ)**
Monomer	Dimer	Monomer
X-Ray (resolution: 2.2 Å)	X-Ray (resolution: 2.3 Å)	X-Ray (resolution: 1.57 Å)
2 Counter ion	2 Counter ion	2 Counter ion
Halobacerium Salinarium (Archae)	*Chlamydomonas reinhardtii* (green alga)	*Acetabularia acetabulum* (green alga)

(b) Templates for Homology Model

Figure 7.1.1: Swiss model homology model results (top) and template proteins (bottom)

positions in all homology models, except of the model based on 3VVK. Here, the binding pocket has no space for the retinal molecule, therefore the alignment of the retinal lead to an abnormal placement of the retinal. This model has been discarded from further analysis. Nevertheless, the accordance of the residue positions in the binding pocket indicates a reliability of the homology models in the binding pocket. Since the focus of this study lies on the interaction between the retinal and the binding pocket, the accuracy of the homology models condensed in the binding pocket is absolute sufficient.

For a better validation of the homology models an additional homology model has been created with the software package phyre2 (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index)[141] and have been compared to the Swissmodel structures. The phyre2 model contains the amino acids in the binding pocket on the same position as the models build with Swissmodel. This agreement between homology models constructed by different methods is a further hint for the reliability of the homology models.

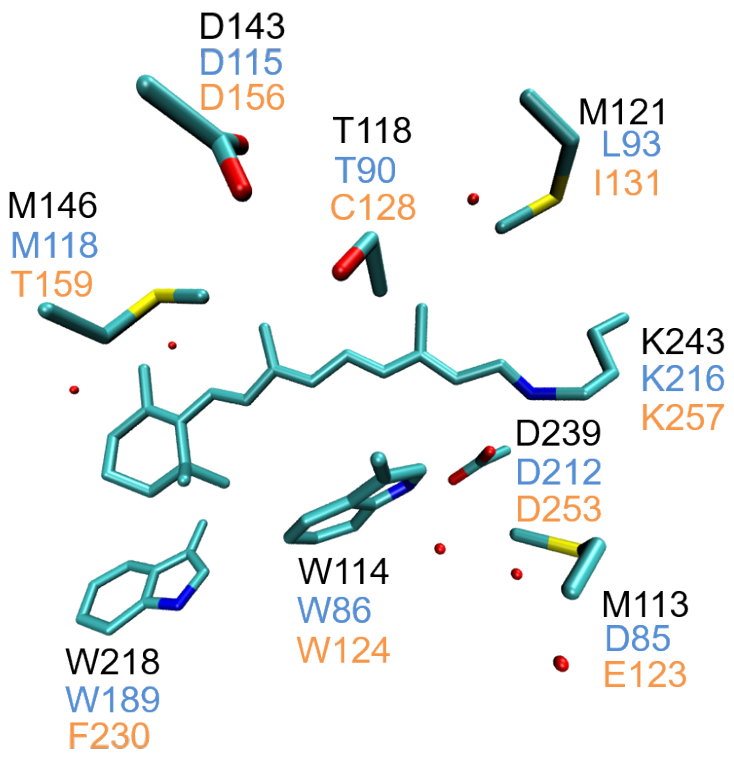


Figure 7.1.2: Binding pocket of the homology models, this residues are at the same position in all models. For comparison, the residues of BR (blue) and of ChR-2 (orange) at similar position relative to the retinal.

7.2 MM equilibration

In order to get water molecules in the homology models, the crystal structure water of the respectively template has been fitted and built in the homology structure. Afterwards, the homology models have been duplicated and combined to a dimer. The dimers have been embedded in a POPC bilayer in a box with bulk solvent similar to the Channelrhodopsin (see chapter 6 and Figure 6.3.1). Then, each box have been equilibrated by MM methods using the GROMACS programm package[110,111] with the Charmm36[107,108] force field.

The first step was the minimization of the potential energy. This was achieved by calculating 10000 steps with the steepest-descent method (tolerance of 1000 kJ mol^{-1} nm^{-1}).

The next step has been the equilibration of the solvent. This was performed by simulating a NVT-ensemble for 2 ns at 300 K. All simulations have been performed with a time step of 2 fs. The heavy atoms of the protein and lipids have been restrained to their initial positions by harmonic potentials (force constant: 1000 kJ mol^{-1} nm^{-1}).

Subsequently the protein and the lipid bilayer are equilibrated at a pressure of 1 bar in the NPT-ensemble over 7 ns, using the Parrinello-Rahman barostat. The position restraints on the protein were still applied to obtain the initial structure for the production (MM- and QM/MM-) simulations.

7.3 QM/MM Simulations and Excitation Energy of Rh-BI state

As it was shown for ChR-2, the interaction in the active site and the description of the hydrogen bonding patterns in the vicinity of the RSBH$^+$ is only described correctly with QM methods. Therefore QM/MM simulations have been performed on the homology models containing the retinal, the covalently linked K243, the counter ion D239 and a nearby water molecule in the QM zone.

The five equilibrated homology models: 3UG9, 5AWZ, 1XIO, 2KSY, 1PXS have been used and 1 ns long QM/MM simulations have been performed. The GROMACS programm package[110,111] with the Charmm36[107,108] force field has been used for the MM part, while the in GROMACS implemented SCC-DFTB3[88,116,117] with the 3ob-parameter[118] has been used to simulate the QM-part. The QM/MM boundary was chosen between C_α and C_β atoms of the amino acid side chain and the standard link atom approach was used. The simulations time step was set to 1 fs.

In order to detect, if there are structural changes within 2 ns simulation time, the RMSD along the trajectory of the three models (5AWZ, 3UG9 and 1XIO) using the MMTSB toolset[142] has

been calculated to get an averaged structure. The RMSD has been calculated between the averaged structure of the whole trajectory and the structure at a certain point of the trajectory. RMSD show after an initial adapting to the structure a constant value over the remaining simulation time (see Figure 7.3.1).

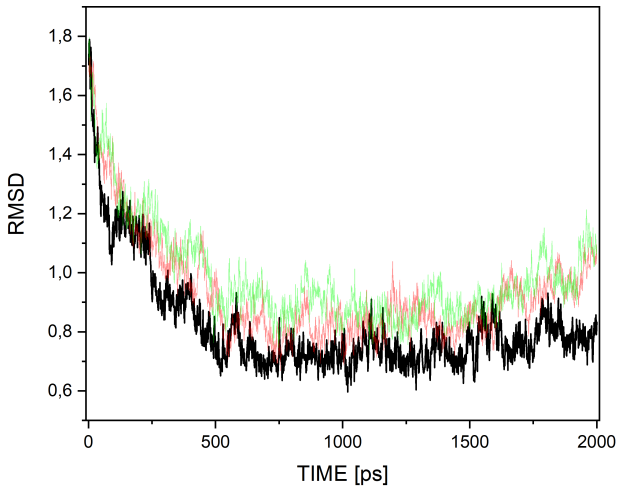


Figure 7.3.1: RMSD of the 1XIO (green), 5AWZ (red) and 3UG9 (black) homology models over 2 ns QM/MM simulation. The RMSD is calculated between an averaged structure over the whole simulation time and the structure at the certain time point in the trajectory.

The absorption of the retinal is very sensible to the electrostatic environment, for example the excitation energy shift between BR and ChR-2 came from different amino acid composition in the binding pocket. In a former study by Welke et al. mutation of the amino acids in the binding pocket of BR to the amino acids of ChR-2 has lead to the correct hypsochromic shift in the excitation energy.[98,99] Therefore, the composition of the binding pocket is crucial to reproduce the correct shift in the excitation energy. That is why the excitation energy is a good indicator for the reliability of the homology model, especially in the binding pocket and therefore is determined on the constructed homology models of HKR1.

The OM2/MRCI method has been tested to give reasonable results with an systematical shift compared to SORCI calculations of the excitation energy in rhodopsins[115,121] and is a fast and efficient method. Furthermore, OM2/MRCI has been shown to reproduce the experimental shifts in former studies between different rhodopsins (see Table 7.3.1).

The excitation energy has been calculated with the OM2/MRCI method on snapshots of every ps along the QM/MM trajectory. The excitation energies are concentrated in histograms and are fitted to Gaussian curve $E = A_1 * exp(A_2 * (A_3 - x)r)$ to determine the maximum of the histogram. The averaged maxima A_3 are summarized in Table 7.3.2.

The experimental absorption maximum of HKR1 is at 2.55 eV (485 nm)[68] and is about 0.37 eV blue shifted compared to BR (2.18 eV or 570 nm). Comparing the calculated

Table 7.3.1: Excitation energy shift compared to BR of different rhodopsins.

Protein	Experimental shift [eV]	Calculated shift by OM2/MRCI
ChR2-WT	0.44	0.55
ChR2-C128T mutant	0.40[115,121]	0.46
ChR2-E123T mutant	0.35	0.36
SRII	0.32[143]	0.30
Rh I	0.29[144]	0.19

excitation energies of the homology models (Table 7.3.2), the experimental shift, relative to BR of 0.37 eV is not reproduced by the homology models, except for the 3UG9 model. Note, that the template for the 3UG9 model is the C1C2 chimera of the Channelrhodopsins from the same organism and genetically affinity between HKR and Channelrhodopsin is probably much higher than for the templates.

Table 7.3.2: Shift of calculated excitation energy of the homology models with respect to BR. The BR excitation energy was chosen from the QM/MM simulations described in chapter 6 with a value of 2.62 eV.

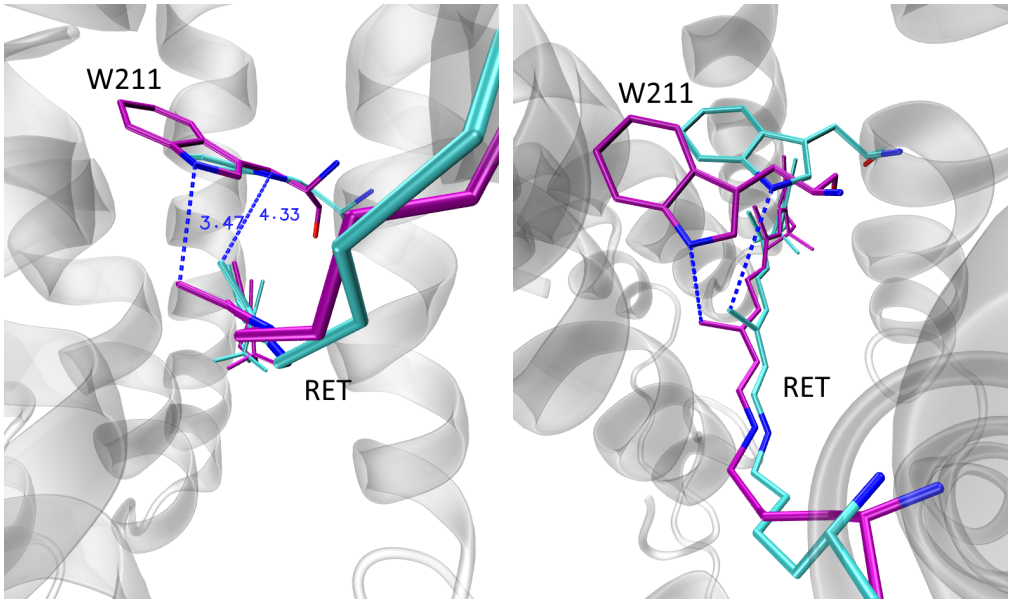
Homology model	Excitation energy shift compared to BR [eV]
Experiment[68]	0.37
1XIO	0.05
5AWZ	0.01
1PXS	0.02
3UG9	0.28
3VVK	-0.38
2KSY	0.14

Nevertheless, the binding pocket of each homology model consist of the same amino acids, except the binding pocket of the homology model of 3VVK. Here the retinal is not aligned in the right position. This leads to the high blue shift of the average excitation energy. This structure model is discarded in the following study.

Interestingly, in the 3UG9 model the excitation energy reaches almost the experimental shift, the excitation energy is more blue shifted compared to the other models, despite that the amino acid composition in the binding pocket is the same in all homology models and is therefore not responsible for the shift of the excitation energy. A first analysis of the binding pocket shows a stronger twisted retinal molecule in the 3UG9 model compared to the other models. A twisted retinal molecule could be responsible for a blue shift of the excitation

energy, which has been shown by Wanko et al.[144] Therefore it is concluded, that the torsion angle of the retinal shifts the excitation energy. Since the other homology models did not show the twist of the retinal molecule, the question arise, why the retinal is not twisted in these models, even though the amino acid composition in the binding pocket is the same in all models. An extended analysis of the binding pocket shows that the C13-methyl group points towards residues D143 and T118 and this is not the case in the other homology models.

Residue W211 pushes it closer to the C13-methyl group in 3UG9 model compared to the other models and therefore is responsible for the twisted retinal (see Figure 7.3.2)



(a) (b)

Figure 7.3.2: Retinal configuration in 3UG9 (purple) and 1XIO (cyan) model

Due to the wrong predicted spectroscopic shift of the models 5AWZ, 1XIO, 2KSY and 1PXS, these models will be discarded in the further investigation and the best structures, 3UG9 has been used for further investigation of the photocycle.

7.4 Long Time MM Simulations of the Homology Models

Previous study on ChR-2 (see chapter 6) demonstrated the importance of a well-equilibrated structure and that homology models can resemble the real protein structure with all their properties. Even the E90 orientation in ChR-2 could be found in the homology model of ChR-2, but the E90 conformation turned up only after extended sampling time. Therefore, the 3UG9 homology model is simulated over a timescale of 300 ns to see if any structural rearrangements occur.

The structure in the binding pocket is conserved, no rearrangements occur within 300 ns

MM-MD simulation (Figure 7.4.1). Thus, the structure seems to be stable and reliable and can be used for further analysis.

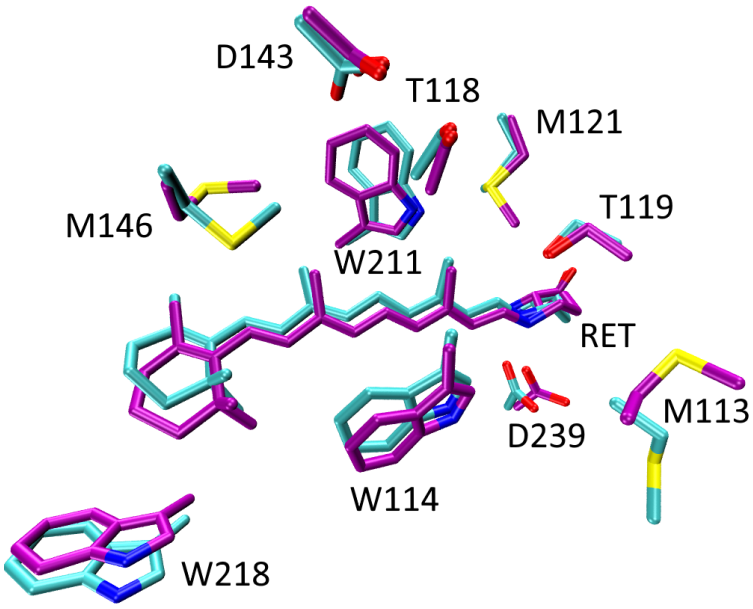


Figure 7.4.1: Binding pocket of 3UG9 homology model before (C-atoms in cyan) and after (C-atoms in purple) 300 ns simulation time.

7.5 Isomerization Pathway

The photoreactions of both stable states in HKR1 are triggered by the light induced photoisomerization of the C13=C14 bond in the retinal molecule. The photodynamics and isomerization mechanism of retinal in rhodopsins is already well investigated in Bacteriorhodopsin (BR), *Anabaena* sensory rhodopsin (ASR) and in Rhodopsin (Rh). Different isomerization mechanisms have been found in the different rhodopsins, for example the double bicycle pedal mechanism in BR, which is a concurrent rotation of the three adjacent double bonds C11=C12, C13=C14, C15=N. However, in all rhodopsins, the protein environment is responsible for these different isomerization mechanism and needs to be included in the calculation. But, this is a challenging task for computational methods, since already the high level *ab initio* methods, which are required for the correct description of the retinal isomerization are complex and computational demanding. They are needed to describe the effect of the electron correlation, which has to be included in the calculation for the correct description of the mechanism. Furthermore, the computationally less costly DFT methods, which include electron correlation as well are not suitable for retinal, since they struggle with the description of the charge transfer character of the electronic transition of retinal. Usually, the *ab initio* QM calculations of the retinal are combined with a MM treatment of the protein environment, in a QM/MM strategy. Whereby, the combination of the CASSCF

method with MM method provide the best compromise between computational cost and balanced description of ground and excited state. The most balanced description of the isomerization would be achieved with the combined CASPT2//CASSCF strategy. Here the geometry is obtained by CASSCF method, while the energy is calculated with the CASPT2 method, which additionally includes the dynamic electron correlation and grant a quantitative picture of the isomerization. However this strategy is more complex and therefore more costly. In this study a qualitative picture is sufficient rather than a quantitative picture of the retinal isomerization mechanism in HKR1, therefore only the CASSCF method has been used for the relaxed scan calculations. For all calculations in this section, the program package MOLCAS 8.2[145,146] and Tinker 6.3[147] have been used and an active space of 12 electrons and 12 orbitals including all π-orbitals has been chosen for the calculations.

In Penzkofer et al.[148] the isomerization mechanism has been investigated by absorption and fluorescence spectroscopy. At the Franck-Condon point an excitation in the S2 state is reported. Afterwards, a relaxation from S2 state to S1 state with a time constant of 0.65 ps occurs with a followed rotation of C13=C14 bond in the S1 state up to the funnel with a internal conversion. This occurs with a timeconstant of 5 ps and based on this constant a barrier of 2kcal/mol is identified along the reaction path in the S1 state. The recovery to S0 state takes place via a twisted internal conversion of the retinal leading to the isomerized product in the ground state.

7.5.1 Optimization of Ground State Geometry

The first step in the photoreaction is the transition in the excited state at the Franck-Condon point. An optimization of the retinal structure in the protein with the CASSCF methods is required to provide a consistency of the isomerization pathway. Therefore, the last snapshot of the 1 ns long QM/MM MD-simulation from the 3UG9 model has been used and transformed in the MOLCAS/Tinker format and has been optimized in the ground state with the CASSCF method and a QM/MM approach. Based on this structures, a QM/MM CASSCF single point energy calculations in the excited state has been performed. Additionally, for comparison, on the last snapshots of the QM/MM MD-simulations of the 5AWZ and 1XIO homology models an optimization and a single point excitation energy calcualtions has been performed as well.

The excitation energy at the Franck-Condon geometry of the 3UG9 model shows a higher oscillator strength in the S2 than in the S1 state. This is in accordance with the experimental results.[148] While, in the 5AWZ and 1XIO models the bright state is the S1 state. This is an another evidence for the reliability of the 3UG0 model.

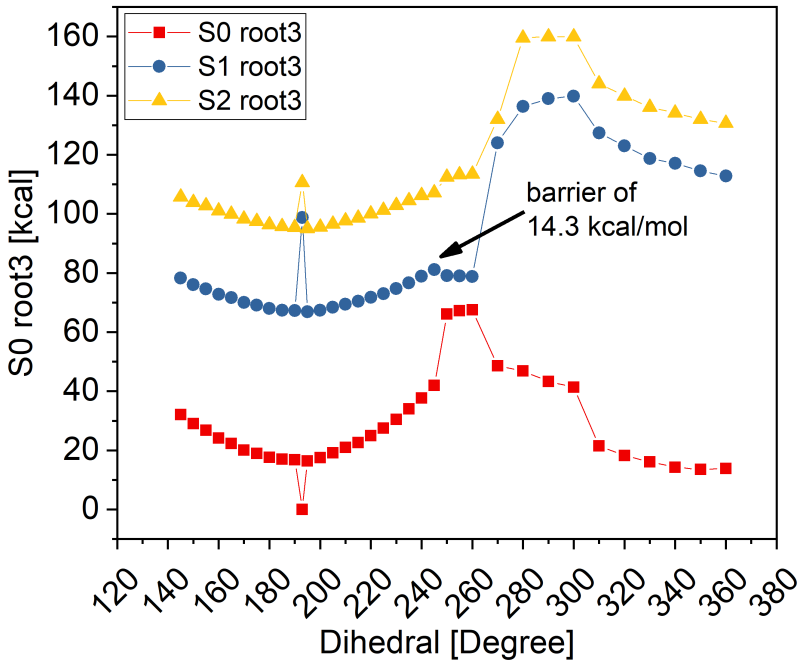
7.5.2 Isomerization Pathways

The isomerization of the C13=C14 bond in retinal has been investigated along the reaction coordinate along the rotation around the C12-C13-C14-C15 dihedral angle in the so-called relax scan. The retinal is optimized with a stepwise constrained rotation around the dihedral angle, while the protein environment has not been relaxed during the optimization. The dihedral angle of C13=C14 at the Franck-Condon point is 193°. Two relaxed scans have been calculated along the reaction coordinate of this dihedral angle in 5-10° steps. One energy profile is along the S2 surface and the other along the S1 surface. The pathway along the S1 surface shows a barrier of 14.8 kcal/mol (see Figure 7.5.1a), while the pathway along the S2 surface shows only a barrier of 5.8 kcal/mol (see Figure 7.5.1b).

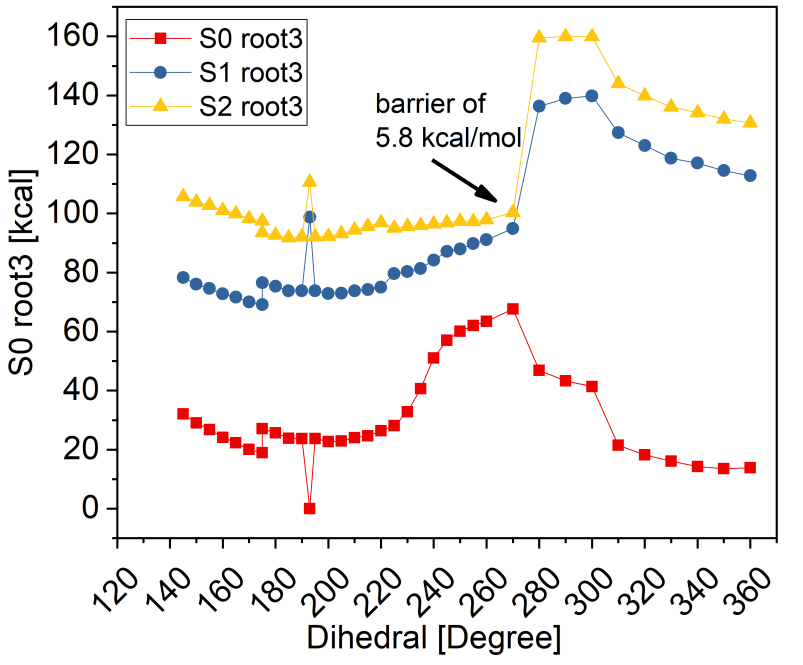
The relaxed scan along the S1 state is not reasonable, since the energy barrier is too high for the experimental time constant of 5 ps. By contrast the isomerization pathway at the S2 surface shows only an energy barrier of 5.8 kcal/mol, but in experiment a fast time constant of about 600 fs has been determined for the first relaxation. Furthermore, the energy gap (about 30 kcal/mol) between the S2 or the S1 and the ground state is to big for a spontaneous relaxation to the ground state. The too high energy barrier compared to the experimental values could result from the lacking consideration of the dynamic electron correlation by the CASSCF method. The lack of the dynamic electron correlation can lead to wrong stabilization of the electronic ground and excited states and provide possible incorrect energies.[149–154] In this case, it could mean that the energy barriers are over estimated with CASSCF. Considering this, the energy profile along the S2 surface could be reasonable for a time constant of 5 ps, but in experiments a second very fast time constant of about 600 fs has been measured. The origin of this time constant could not be explained by these calculations.

The combination of the two energy profiles from above provide a possible isomerization pathway with a longer relaxation from the S2 state in the S1 state and a fast relaxation from the S1 in the ground state (see Figure 7.5.2). The dihedral angle of C13=C14 of the retinal in the S2 state is twisted up to 245° and when it has to over come in this state only a barrier of about 4.9 kcal/mol, which is reasonable compared to the experimental results. Than it relaxes to the S1 state where no barrier is present before at 270° it relaxes back to the ground state. Here, the dihedral angle rotates further up to the ground state minimum at 350°. During the isomerization the $RSBH^+$ points to the counter ion D239. Until the retinal is relaxed back to the ground state and the dihedral angle of C13=C14 is about 300° the $RSBH^+$ reorients and points to the extracellular side of the protein (see 7.6.1b).

However, in these models the relaxation of the protein is not considered. The protein is fixed during the optimization. The relaxation of the protein environment could lead to an further decrease of the energy barriers. Furthermore, the strategy, which is used in this



(a)



(b)

Figure 7.5.1: Energy path of the rotation of the C13=C14 bond in the excited state of 3UG9 homology model. a) optimized along S1 surface, b) optimized along S2 surface

study is problematic since there is a bias on the retinal molecule, due to the restraints on the dihedral angle of the C13-C14 bond. An excited state dynamic simulation would solve this problem, but the methods used here, are computationally too expensive to propagate a dynamic simulation in the excited state. Semi-empirical methods would help to over come the computational issue. Unfortunately, there are not many semi-empirical methods, which give a reasonable description of the excited state surface of the retinal. Nonetheless, the semi-empirical methods needs to undergo extensive test procedures and requires a validation of the application. One of these promising methods is OM2/MRCI. A detailed benchmark was also part of this work and is described in the next chapter (chapter 8).

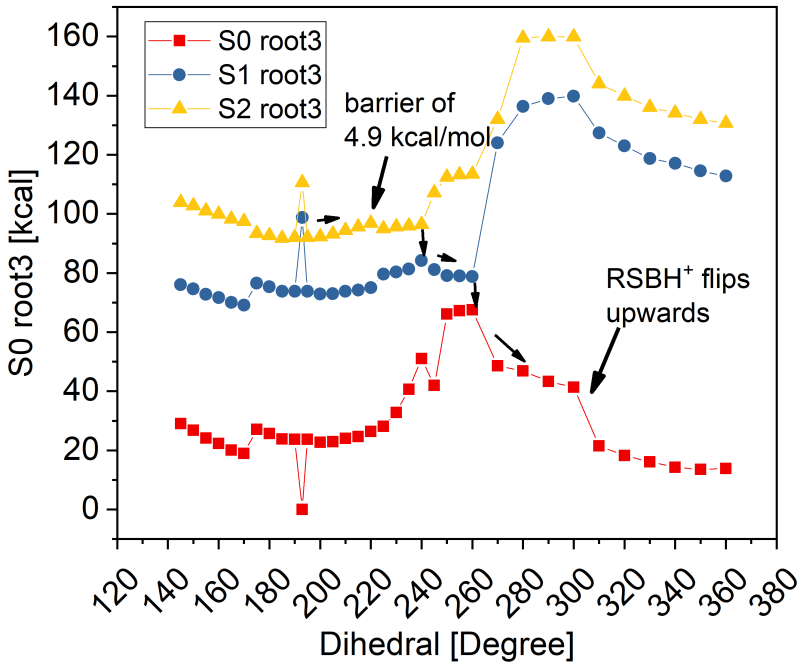
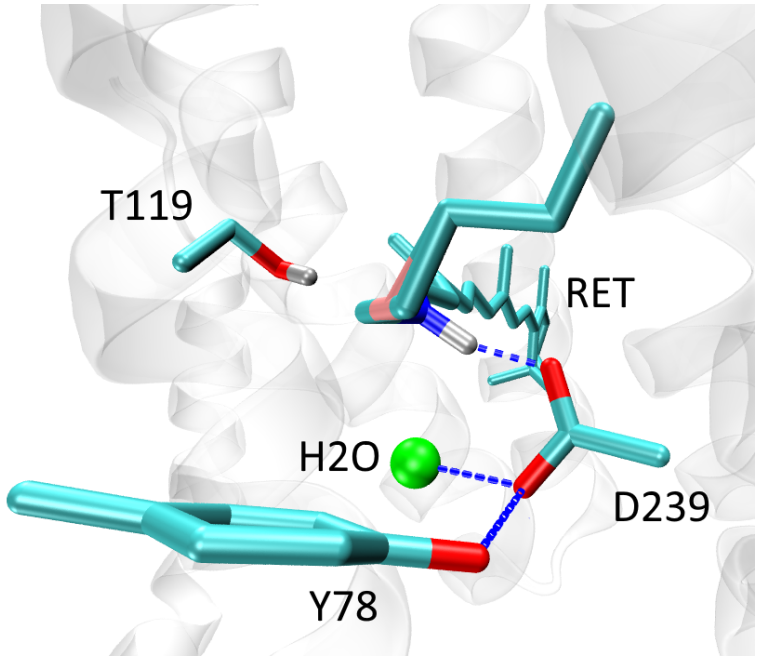
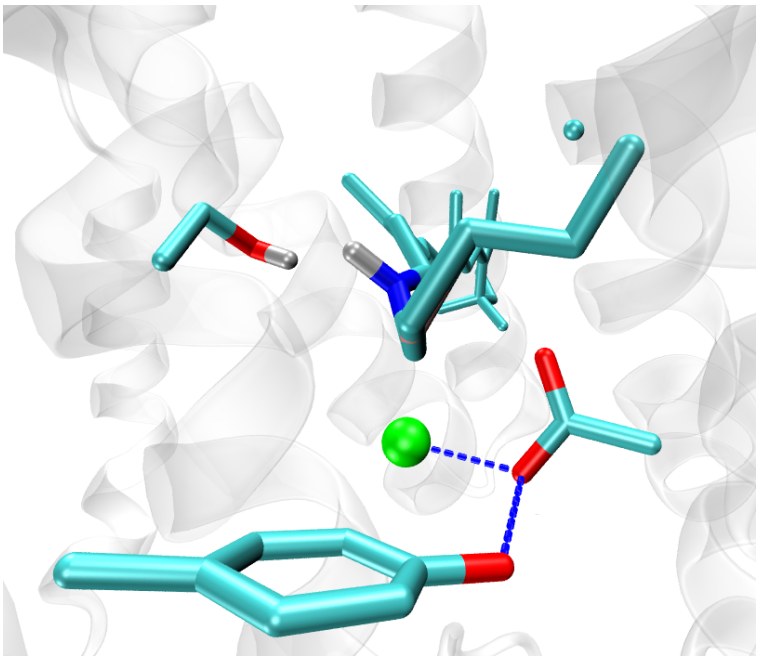


Figure 7.5.2: Proposed photoreaction pathway of C13=C14 isomerization of retinal in HKR1. The pathway is a combination of the optimized rotations in the S2 and S1 state from Figure 7.5.1. The propagation starts in the S2 state with a barrier of 4.9 kcal/mol and a relaxation to the S1 state before the relaxation in the S0 state.

The energy profile for the isomerization does not reproduce the experimentally determined time constants. However, the main focus in this project lies not on the investigation of the detailed mechanism of the retinal isomerization, but rather in the determination of a reasonable 13-*cis*, 15-*syn* retinal structure in the binding pocket for the investigation of the further intermediates of the photocycle. Therefore, the 13-*cis*, 15-*syn* structure with the minimum energy has been chosen for further analysis.



(a)



(b)

Figure 7.6.1: $RSBH^+$ orientation in (a) optimized ground state structure at the Franck-Condon point containing a all-*trans* retinal and (b) optimized ground state structure after the isomerization containing 13-*cis*, 15-*syn* retinal and a C13=C14 dihedral angle of 350°.

7.6 13-*cis*, 15-*syn* Retinal in HKR1

The minimum structure with a dihedral angle of 350° (see Figure 7.5.2) has been used as starting structure for a QM/MM MD simulation in the ground state. Therefore the same setup as described in section 7.3 has been used.

The starting structure with 13-*cis*, 15-*syn* conformation of the retinal in the HKR1 shows that, the $RSBH^+$ does not have a permanent binding partner. It points to the extracellular side (see Figure 7.6.1b).

In a QM/MM simulation of the structure are no structural changes observable over a timescale of 100 ps. The $RSBH^+$ has no hydrogen bonding partner, but some times forms a hydrogen bond with THR122.

The structural rearrangements, which happen from the intermediate INT1 to the intermediate P550, are in a timescale of 453 ps (see Figure 2.2.1) in the photocycle. Therefore a longer QM/MM MD simulation is necessary to see structural rearrangements in the homology model. This is part of the furture work.

Nevertheless, the isomerization lead to a reasonable structure and based on this structure the further investigation of the second stable state is much easier, since the structural properties in the binding pocket are clarified.

7.7 Concluding Remarks and Outlook

Several homology model has been built from different templates and by two different methods. All model show structural similarities in the binding pocket and this indicates the reliability of the homology models. Furthermore, the homology models have been rejected based on the spectroscopic property of the retinal. The best matching homology has been further validated through the excitation energy and the stability over a extended simulation time has been tested.

It has been found out, that the blue shift of the absorption spectrum of HKR1 compared to BR results not from the amino acid composition in the binding pocket, rather from the conformation of retinal. Only if the retinal is twisted in the binding pocket, the blue shift of the excitation energy can be reproduced. Based on this model, the optical properties had been investigated and are in good agreement with the experiment, since the absorption in the S2 state is reproduced by the homology model. This special absorption property of HKR-1 (absorption in S2) is due to the twist of the retinal and could not be observed for the remaining homology models with a non-twisted retinal configuration.

However, the analysis of the photochemical reaction could not lead to satisfying results, since the experimental time constants are not reproduced by the calculated energy

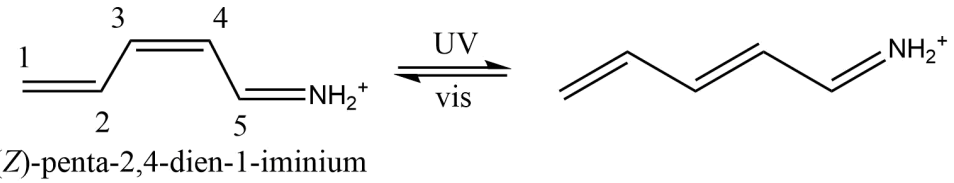
profiles. Therefore, more sophisticated methods are required, which include dynamic electron correlation or enable non-biased dynamic simulations on the retinal in the protein environment. An adequate semi-empirical method, which enables excited state dynamics in a feasible time, is OM2/MRCI and its benchmark is described in the next chapter (chapter 8).

Nonetheless a reasonable isomerized 13-*cis*, 15-*syn* retinal structure is modeled and is suitable for the investigation of the further intermediates in the photocycle of the HKR1. For the future a longer QM/MM simulation needs to be performed to investigate the structural relaxations, which happen because of the 13-*cis*, 15-*syn* configuration of the retinal in the binding pocket. In order to concerning a very interesting question, why the deprotonated RH-UV state is stable over hours, while in other bimodal rhodopsins this state is only stable up to a few minutes? An insight could give a manually deprotonation of the Schiff base followed by a QM/MM simulation in order to gain structural changes and characteristics of this state. This study pave the way for the investigation of the second stable state of the photocycle, since a reasonable structure is crucial and leads to correct and reasonable conclusions in the further research on the intermediates.

8 | Benchmark OM2/MRCI for Surface Hopping Dynamics on Retinal Models

The photoisomerization of retinal protonated Schiff bases (PSBs) is the primary reaction of the rhodopsins and play a key role in various processes occurring in nature. This reaction triggers conformational changes within the protein, which leads over several intermediates to the specific function of the retinal proteins, like the vision process in rhodopsins, proton pumping in Bacteriorhodopsin,[155] channel-opening in Channelrhodopsins[60,62,154] (see chapter 6) or signal transmission in sensory rhodopsins II[52,156] or Histidin Kinase Rhodopsin[68] (see chapter 7). Therefore, it is not surprising that this process has been in focus of experimental and computational research for several decades. On the experimental side the process can be studied by ultrafast time resolved spectroscopy, but computationally, the reaction can be tracked using non-adiabatic molecular dynamics. An accurate simulation of the molecular dynamics requires the description of the ground and excited state on equal footing, in addition the transition through a conical intersection has to be accounted for. The complete active space self-consistent field (CASSCF) method fulfills these requirements and has been successfully applied to various retinal model systems.[157–173] However, CASSCF or similar methods do not consider the impact of dynamic electronic correlation. For spectroscopic calculations complete active space perturbation theory of second order (CASPT2) has been used to reproduce excitation energies. It accounts for the effect of the dynamic electron correlation and this can lead to different stabilization of the ground and excited state and therefore to qualitatively different potential energy surfaces.[149–154] Mixed methods fail in describing the energy surface of protonated Schiff Base model (PSB3).[174] PSB3 (see Figure 8.0.1) is a well studied for the investigation of the photo dynamics of retinal.

In all these studies, the static methods like “relaxed scan”, “Minimum energy pathway” or minimum energy CI” had been calculated and compared between CASSCF and CASPT2. However, the dynamical effect and kinetic energy is neglected in these studies. Liu et al.[175] compared AIMS non-adiabatic dynamic simulations of PSB3 with CASSCF and CASPT2 and observed differences in the efficiency and the torsion coordinate between this two methods. This shows the importance of dynamic electronic correlation in the reaction pathways of



(*Z*)-penta-2,4-dien-1-iminium

Figure 8.0.1: Protonated Schiff base model containing three double bonds is a well studied retinal model, which acts as test case for several computational methods.

protonated Schiff base models. However, CASPT2 is highly computational expensive and a sufficient manner of dynamic calculation on PSB3 had not been feasible yet. A new implementation of XMS-CASPT2 in Bagel program package[176–182] is very efficient due to the availability of analytical gradients and enables to run multiple excited state trajectories, but is still limited to small systems and requires long computing time. However, in order to obtain a statistically meaningful number of simulations hundreds of trajectories are required. Such a large number of calculations is prohibitive. A computationally favorable alternative is the semi-empirical orthogonalization model 2 with multireference configurational interaction (OM2/MRCI) method. It was already tested to give reliable results for absorption energies for bacteriorhodopsin[144] sensory rhodopsin II[143] and Channelrhodopsin (chapter 6). Moreover, surface hopping dynamics with OM2/MRCI show promising results for PSB3,[151] but was not yet systematically tested against the high level *ab initio* methods like CASPT2. This is the main purpose of the present study that compares surface hopping dynamic simulations of XMS-CASPT2, CASSCF and OM2/MRCI for PSB3. The isomerization mechanism will be investigated, in particular the reaction coordinate, excited state lifetime, quantum yield and compared to results of standard CASSCF dynamics. This project is in collaboration with Veniamin Borin and Igor Schapiro at the Hebrew University in Jerusalem. They performed the CASPT2 dynamics, while the CASSCF and OM2/MRCI dynamcis has been performed at the KIT.

8.1 Computational Details

8.1.1 OM2/MRCI dynamics

For the semi-empirical calculations the OM2/MRCI method (orthogonalization model 2 and multireference configuration interaction)[91,172,183] implemented in the MNDO program package[184] has been used. For generating the initial conditions for the excited state surface hopping dynamics, a ground state trajectory over 20 ps with a temperature of 300 K and a time step of 1 fs without velocity scaling has been performed. The total energy is conserved over the ground state trajectory. 100 snapshots of geometries and velocities of every 200

steps of the trajectory have been chosen as initial conditions for excited state dynamics. The dynamics in the excited state have been tested by several setups, since the correct description of the S_1 state and the photo dynamics requires the inclusion of all π-orbitals in the active space. In a previous study, the ROHF was suggested to give the reproducible excited state life times for the PSB3 retinal model.[185] However, the total energy was not conserved with the options mentioned in the study by Keal et al. which indicates a sudden change in the electronic structure. The implemented adaptive time step option solved this problem and ensured a conserved energy of the system.[186] An energy change of less than 0,01% , an orbital overlap of greater than 0.97 and a maximal number of iterations of 10 have been chosen as criteria for the adaptive time step option.

For the SCF calculations, the restricted Hartree-Fock (RHF) method with an active space of 3 occupied and 3 unoccupied orbitals and 6 electrons have been used. All single and double excitations have been taken into account for MRCI.

For configuration interaction procedure, a full CI calculation within the active space (nciref=0) has been chosen. While, the MOs for the SCF calculation have been selected automatically as it is suggested by Keal et al.[151] The threshold for the *p* populations of the orbitals was set 0.4 (default). To ensure that all relevant π-orbitals are in the active space during the dynamic simulation, the orbital tracking method (imomap) has been used. This setup is named in the further writing as AS6. Additionally, calculations have been set up, where the size of the active space can vary. This setup is named AS+.

The Tully algorithm has been chosen to determine the surface hopping during the simulations.

8.1.2 CASSCF Dynamics

For comparison, 100 CASSCF excited state dynamics have been performed with the program package MOLCAS 8.2.[145,146] An active space of 6 orbitals with 6 electrons has been chosen and the time step was set to 1 fs. The total simulation time was 200 fs for each trajectory.

The initial conditions for the excited state dynamics were generated from a RHF Born-Oppenheimer dynamics simulation in the ground state with a time step of 1 fs, therefore the Gaussian program package[187] has been used. The ensemble temperature has been set to 300 K.

8.1.3 CASPT2

Since CASSCF is not the best reference for the excited state dynamics, due to the lack of the dynamic electron correlation effects, additional 200 excited state trajectories with

CASPT2 have been performed and analyzed. This work was done by Veniamin Borin at the Hebrew University of Jerusalem. The XMS-CASPT2 molecular dynamics has been performed using Newton-X program package, v 1.4 developer version, interfaced with BAGEL v1.1 program. The active space consist of six π-orbitals (three bonding and three antibonding). For XMS-CASPT2, the SS-SR contraction scheme was employed and the vertical shift was set to be 0.2 a.u. The cc-pVDZ and cc-pVDZ-jkfit basis sets were used during the computation. The initial conditions (geometries and velocities) were generated by running 20 ps XMS-SS-CASPT2 ground state trajectory, started at S0 equilibrium geometry and randomly generated velocities (at 300 K). Every 100 fs the geometry and velocities were used to initiate the XMS-SA2-CASPT2 excited state molecular dynamics with FSSH algorithm for S1$\rightarrow$S0 transition. For all trajectories the timestep of 1 fs was used.

8.2 Results and Discussion

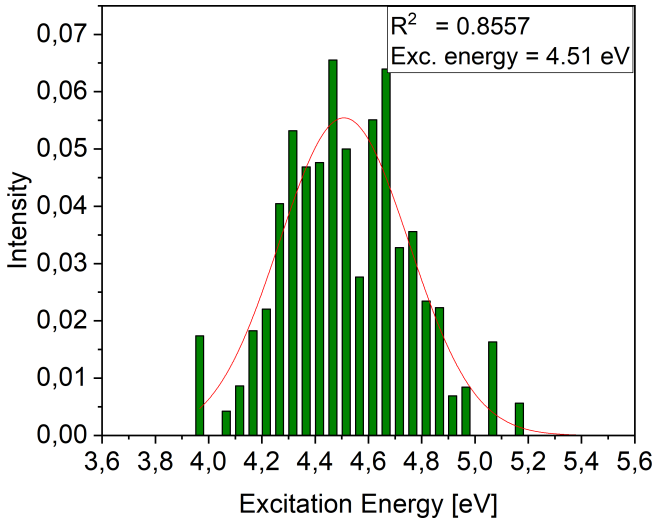
8.2.1 Excitation Energy

Along the ground state trajectories, the excitation energy has been calculated with the corresponding method (Figure 8.2.1), in order to get an insight into the equilibrium state of the ensemble and their distribution at the excited state surface. The excitation energy of the OM2/MRCI ground state trajectory (8.2.1b) and the CASPT2 gound state trajectory (see 8.2.1c) are very similar and the excitation energy maxima are almost equal. 3.98 eV with the OM2/MRCI method and 3.96 eV with the CASPT2 method. The maxima has been determined by fitting the data on the Gaussian $g(x) = a_0 e^{a_2(x-a_1)^2)}$. This is a first indication, that OM2/MRCI could be a reasonable method in describing excited state dynamics of PSB models. Both spectra are narrowed in comparison with CASSCF. Furthermore, the maximum with CASSCF is blue shifted about 0.5 eV, which is in agreement to a previous study.[151] In non of the spectra are sub-peaks observed, which indicates a well-equilibrated ground state trajectory and thermal sampling of the ground state minimum.

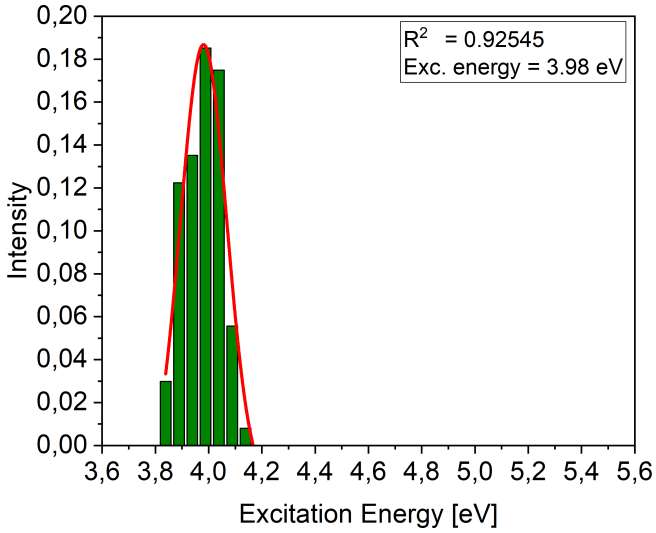
8.2.2 Excited State Life Time and Isomerization Mechanism

One of the important properties, which may be obtained from the excited state dynamics simulations, is the excited state life-time. It can be easily compared with experimental values and different computational methods. Furthermore, it provides the evidence, whether the description of the potential surface in the excited state is reasonable or not, since small barriers or minima affect the dynamic behavior of the molecules in the excited states and results in different exponential decays with probably different decay times.

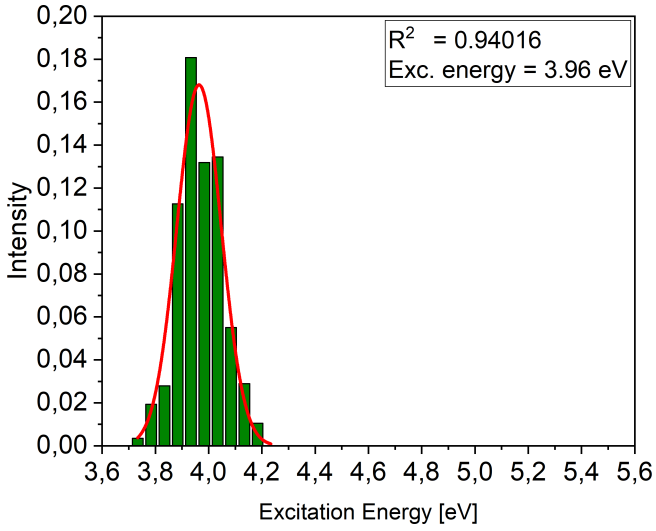
Two sets of semi-empirical excited state surface hopping dynamics with OM2/MRCI and one set of *ab initio* CASSCF excited state dynamics have been performed. For the OM2/MRCI



(a)



(b)



(c)

Figure 8.2.1: Excitation energies of the snapshots from the ground state trajectory from a) RHF-dynamics, b) OM2-dynamics and c) CASPT2 dynamics.

dynamics the orbitals during the trajectory has been tracked automatically based on their π character of the orbital. The two set of OM2/MRCI dynamics differ in the size of the excited state. One set has a fixed active space with 6 orbitals and 6 electrons (AS6). The second set starts with 6 electrons in 6 orbitals, but further inactive orbitals can be included in the active space (AS+). This avoids unphysical gradients due to the mixing of π orbitals at a non-planar PSB3 configuration. Here, all orbitals, showing π character above a certain threshold are added to the active space and are considered in the calculation.

The time for the back transition to the ground state has been determined for the 100 trajectories and the data has been fitted to an exponential decay function: $y = A_1 \cdot exp(-x/t_1) + y_0$ (Figure 8.2.2). The obtained decay life time is for the AS6 trajectories about 735 fs, while for AS+ about 633 fs. The half life $\tau_{1/2}$ ($\tau_{1/2} = t_1 \cdot ln(2)$) account for AS6 461 fs and for AS+ 407 fs.

Comparing these results with decays determined by the *ab initio* methods CASSCF and CASPT2, the OM2/MRCI has a remarkably longer excited state life time. But the half life for the decay is similar to the one of the CASPT2 (see Table 8.2.1). The CASPT2 excited state surface has been found in previous studies by "Minimum Energy Path" calculations to have an energy plateau around the Franck-Condon point of the *cis* configuration of PSB3 along the reaction coordinate of the C4=C5 dihedral angle.[151,175] While, the CASSCF method, shows no barrier along the same reaction coordinate.[151,165,175] This difference in potential energy surfaces could be one reason for the longer excited state life time gained with the CASPT2 method compared to CASSCF. However, in another study,[154] the excited state surface at the CASPT2 level was found to have two different local minima, when optimizing in the first excited state. The corresponding geometries differ in the character of their bond length. One local minimum geometry has inverted bond lengths and the other geometry has not. At which, the inverted bond length minima can only been reached when the optimization is already started from a pre-inverted structure. This structural heterogeneity on the excited state surface has not been observed with the CASSCF method, only the bond-inverted minimum is present. This results arise due to the neglection by the dynamic electron correlation in the CASSCF method. The two equilibrium state geometries have been also reported for the PSB4 and PSB5 models optimized with OM2/MRCI. However, for the PSB3 models only one excited state minimum geometry with non-inverted bond lengths has been reproted.[154]

The relaxed PES scans along the reaction coordinate for the isometrization of PSB5 showed no barrier for the inverted geometry and a small barrier of about 0.5 eV starting from the non-inverted structure with the CASPT2 and the OM2/MRCI method.[154] This, helps to interpret the different excited state life times of the CASSCF, the CASPT2 and the OM2/MRCI dynamic simulations in the present work. The thermal sampling of the ground state provides a balanced distribution of geometries along the ground state minimum. These distribution

has been used as starting structures for the excited state dynamics. Therefore, the starting geometries for the excited state dynamics vary in bond lengths, angles and torsional angles. The evaluation of the excited state dynamics is therefore on a statistical set of different geometries.

The CASSCF life-time is the shortest, since with CASSCF no barrier in the excited state has been reported[151,165,175] and therefore the propagation along the reaction coordinate is not hindered, whereas the CASPT2 dynamics show a longer excited state life time. This results from the fact that, the excitation from the geometries with non-inverted character in the ground state leads to the non-inverted minimum in excited state, and starting from this minimum a barrier has been reproted in Dokukina et al.[154] and the propagation along the energy surface needs to overcome a energy barrier. While the geometries with an pre-inverted character in the ground state, showed in the previous study[154] no barrier, which leads to the barrier less propagation. Considering all trajectories, the determined total excited state life time is longer with CASPT2 than with CASSCF.

The still longer excited state life time with the OM2/MRCI method can be explained, by the fact, that there is only one non-inverted excited state minimum. For the non-inverted geometry in Dokukina et al. a barrier has been reported along the reaction coordinate with OM2/MRCI.[154] Therefore, all trajectories needs to overcome a barrier while propagating along the energy surface in the excited state and this delayed the relaxation back to the ground state of all trajectories. And the total excited state life time is longer.

The isomerization mechanism evaluated in the dynamic simulations, shows that in OM2/MRCI and CASPT2 the isomerization takes places mainly about the C3=C4 double bond, while in CASSCF an additional isomerization of the N=C5 double bond occurs. The results for the isomerization mechanism gained with OM2/MRCI correspond to them gained with CASPT2. OM2/MRCI is in good agreement and resembles the CASPT2 mechanism.

However, the quantum yield is underestimated with the OM2/MRCI method, while CASSCF and CASPT2 quantum yield are similar to each other.

8.3 Concluding Remarks

Comparing CASPT2 and CASSCF results, the discussion shows the importance of the dynamic electron correlation for the description of the excited state properties as well as for the prediction of the isomerization pathways. While the benchmarking of the OM2/MRCI method for PSB3 model system, gives reasonable results. The OM2/MRCI method is able to reproduce the main mechanism, but the CASPT2 excited state life-times could not be reproduced. OM2/MRCI overestimates the barrier for the PSB3 excited state surface. However, promising calculations have been done with the PSB5 model,[154] where the two minima reported at the CASPT2 level are also present at the OM2/MRCI method. Therefore,

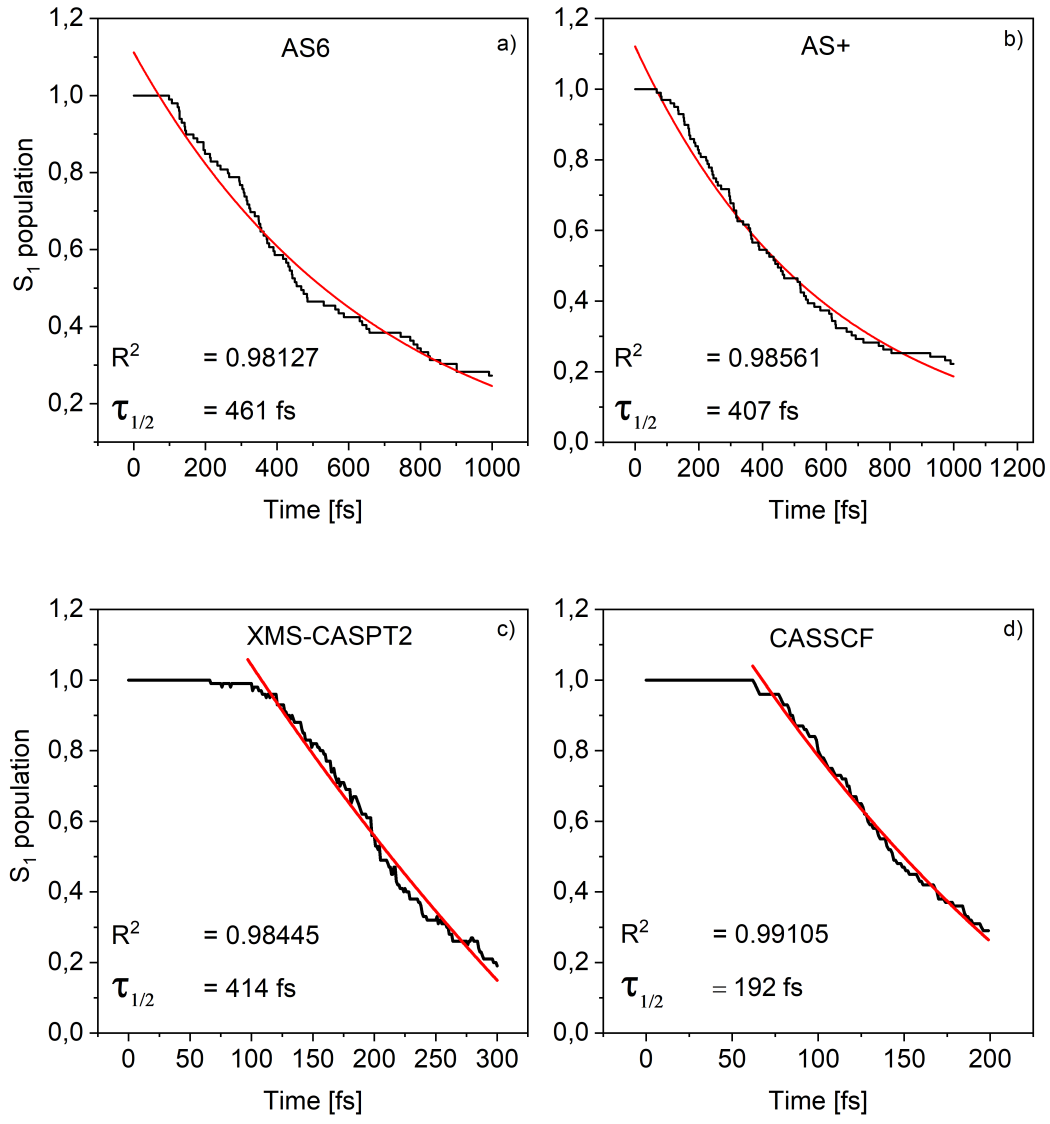


Figure 8.2.2: Excited state population decay of 100 trajectories propagated with different setup. The half life of the decay is shown in the figure $\tau_{1/2}$.

Table 8.2.1: Exponential fitting parameters and quantum yield analysis for the OM2/MRCI AS6, AS+, CASSCF and CASPT2 trajectories.

	OM2/MRCI		CASSCF	CASPT2
Fitting parameters	AS6	AS+		
half life	461	407	192	414
S1 life time	735	633	176	244
y0	-0,00198	-0,02228	-0,94816	-2,09637
A1	1,11	1,14	2,48	3,71
t1	665	588	277	597
k	0,0015	0,0017	0,00361	0,00167
total no. of trajectories	97	96	100	200
Product formed over C3=C4 isomerization	38	34	44	93
Product formed over N=C5 isomerization	0	0	2	0
Return to parent state	30	32	4	84
Return to S0	27	27	16	192
Quantum yield	0,39	0,35	0,44	0,46

the dynamic simulations in the excited state of longer PSB models could lead to the promising results and should be tested for further research.

9 | Describing Fluorescent Behavior of Flugi-2 Molecule with Classical Force Fields

The following project, described in this paper is a collaboration with the group of Tomasz Wesolowksi at the University in Geneva. We performed the parametrization and the trajectories analysis, while the group of Tomasz Wesolowksi performed the frozen-density embedding calculations.

Flugi-2 (Figure 9.0.1) is an environmentally sensitive fluorescent dye, which has been synthesized first by a combinatorial chemistry approach.[188] Within this study, Flugi-2 has been found to give the highest absorption and emission intensity of all the products in the synthesis. Therefore, it has been selected as test case for this project. Furthermore, the measurement of the absorption and the emission spectrum in different solvents shows, that a solvatochromic shift occurs only in the emission spectrum, but is negligible in the absorption spectrum. Thus, the large experimentally measured solvent effect on the Stokes shift can be explained only by the solvatochromism in emission. The largest Stokes shift was observed in DMSO (Figure 9.0.2).

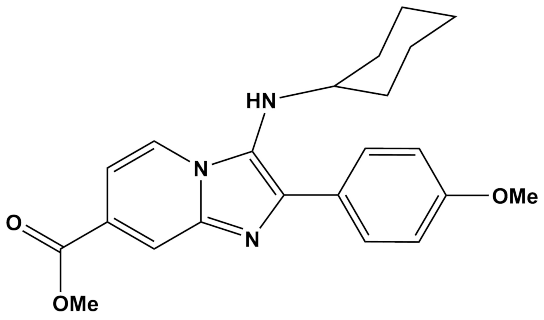


Figure 9.0.1: Structure of the Flugi-2 molecule

The shift of the fluorescence wavelength could result from the increased dipole moment of the molecule in the excited state. But the detailed mechanism, how the solvent interacts with the molecule and therefore shifts the fluorescence wavelength is not known, yet. The main

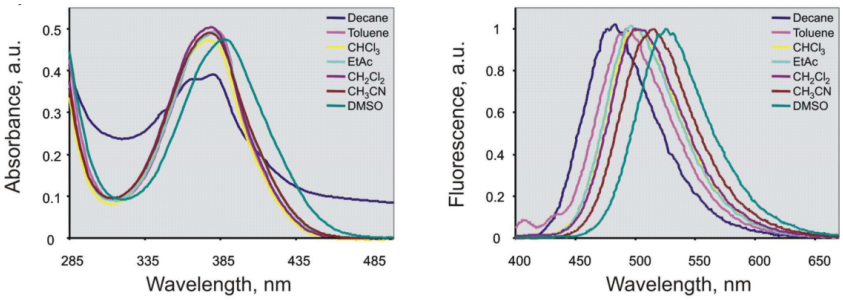


Figure 9.0.2: Absorption and emission spectrum of Flugi-2 molecule in DMSO and other solvent[188]

topic of this chapter is the investigation of the excited state of the Flugi-2 molecule and the interaction between solvent and chromophore.

The solvatochromism effect is already difficult to describe in the ground state, because electrostatic, dynamic and polarization effects of the surrounding solvent molecules have to be considered in the calculations. Furthermore, hydrogen bonds and geometrical reorientation of the chromophore is crucial to understand the shift of the excitation or the emission energy. Usually, QM/MM calculations are applied to describe the interaction between chromophore and solvent. However, the common QM/MM approach neglects the polarization effects between the chromophore and the solvent molecules (more details see section 5.4). Thus, one has to switch to an alternative approach, for example the frozen-density embedding method. The description of the solvent as an electron density and the consideration of this electron density in the Chromophore Hamiltonian, includes the electrostatic interaction between chromophore and solvent in the excitation energy.

Beside the above mentioned difficulties, the description of the solvatochromism in the emission spectrum of the Flugi-2 molecule requires a dynamic simulation in the excited state. Unfortunately, standard quantum mechanical methods are not able to calculate dynamic simulations in the excited state with an explicit solvent environment.

Furthermore, the frozen-density-embedding approach can only calculate excitation energies and their solvent contribution on a given geometry. Therefore, geometries of the chromophore and the solvent molecules are still necessary and have to be provided for the calculations with the frozen-embedding method. Thus, the geometries have to be obtained by classical MD simulations. Our part of the project was to characterize the structural properties and parametrize the Flugi-2 molecule for classical force fields and afterwards, perform the classical MD simulations of the excited chromophore with the surrounding solvent (DMSO) in a thermodynamically equilibrated system. Since the geometry of the chromophore is allowed to fluctuate around the equilibrium excited-state geometry, the dynamical and vibrational oscillations of the molecule in the ground or rather in the excited state can be imitated by the simulations. The usage of classical MD simulations for the description of a chromophore in the excited state embedded in a solvent is a totally new

approach. Thus, if the trajectories are able to reproduce the Stokes shift and the distribution of the experimental fluorescence spectrum due to the DMSO solvent, detailed analysis of the structure and the excitation energy is necessary. Therefore, the energy gap between excited state and ground state was calculated by quantum mechanical methods on snapshots of the trajectory.

In this chapter, the structural variation from the ground to the excited state have been analyzed as well as correlations between the structural pattern and the excitation energy have been found. Furthermore, it could be concluded, that implicit solvent models are not able to describe the interaction between chromophore and solvent correct, which leads to an underestimation of the solvent shift in the Flugi-2 molecule.

9.1 Computational Methods

Geometry Optimization and Single-Point Energies

The parametrization requires a well-optimized geometry in the ground and accordingly in the excited state, because the optimized structures are the basis to determine the parameters for the force field, which describe the behavior of the molecule in MD simulations. Thus, the Flugi-2 molecule was optimized with several quantum mechanical methods and the excitation energies was compared with the experimental values to find out which method can reproduce the experimental Stokes shift. The results and further details are shown in the appendix Table 1.

The high level *ab initio* CC2 method[189–191] reproduces the experimental absorption and fluorescence in the decane solvent (see appendix Table 1). Because decane is a nonpolar solvent and mimics the gas phase environment. Before the Flugi-2 geometries of the ground and excited state with the CC2 method were optimized in gas phase. All CC2 calculations were done with the program package Turbomole Version 7.0[192] and the def2-TZVP[193] basis set. The bond lengths of the π-system (imidazopyridine and benzene) and the dihedral angle between the imidazopyridine and the benzene ring were compared with experimental results. It can be concluded, that the optimized CC2 structure agree well with the in the literature reported crystal structure.[194]

The optimizations with DFT/ωB97X and the 6-31G* basis set were done with the Gaussian09 program package.[187] For all other optimizations and single point energy calculations with DFT, the ORCA program package[195] and the def2-SV(P) basis set were used.

Parametrization

The parametrization was done with AmberTools15[196] and the standard General Amber force field (GAFF)[197] on the optimized CC2 structures. The standard procedure of AmberTools15 for the parametrization was used. The bond lengths and the force constants of the Flugi-2 molecule have to be modified after the parametrization, because the standard parameters determined by the parametrization procedure of the GAFF did not match to the bond lengths reached by CC2 minimization.

In Figure 9.1.1 all force constants of the GAFF-force field are plotted against the bond lengths of C-C bonds, the bond character (single, double bond etc.) is neglected. The plot is linear in the region 1.3 to 1.4 Å. Thus, the force constants of the bonds have been determined basedon the bond lengths in the CC2 structure.

In order to test the parametrization, a minimization with the parameters has been performed and the bond lengths were measured. Based on the difference to the minimized CC2 reference structure, the parameters have been changed. Using this approach the parametrization has been iteratively improved.

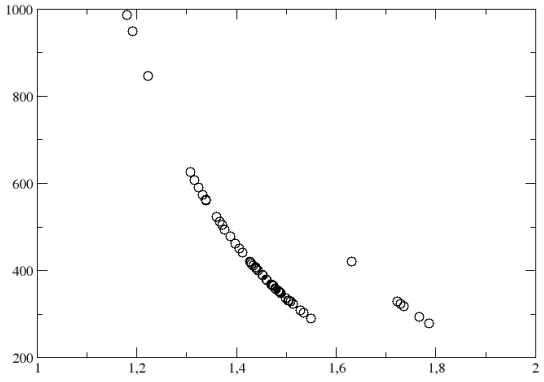


Figure 9.1.1: Force constants for the C-C bond lengths of General Amber force field (gaff).

Charges

The charges for the ground state parametrization were determined by the standard approach with a Hartree-Fock calculation and a Merz-Singh-Kollman-population analysis and the RESP-Fit.[198] However, this procedure is only for molecules in the ground state. The charges for the excited state have to be found with an alternative approach.

It is not possible to get the electron density from a CC2 calculation, thus the electron density

from a DFT calculation with CAM-B3LYP functionals was used and fitted to the atoms. The DFT method with range-separated functionals was used, because these functionals behave better for excitations with charge transfer character. The ωB97X range-separated functional was tested for the excitation energies and the structure optimization of the Flugi-2 molecule (see in appendix Table 1).

The charges for the excited state have been received with a TD-DFT calculation with a related functional, CAM-B3LYP,[199] in the first excited state by using the Koala program package.[200,201] This was done by Sebastian Höfener. The electron density was fitted with an ESP Fit on 14 points per atom and a distance of twice the VdW-radius.

Dihedral Angle C8-C10

The dihedral between the imidazopyridine and the benzene ring (atom labels see Figure 9.2.1) is not included in the GAFF. Thus, parameters for this dihedral have to be determined. The potential of this dihedral is obtained by fixing the dihedral every 10° and optimizing the Flugi-2 with the fixed angle in the ground and in the first excited state with the CC2 method with the program package Turbomole Version 7.0.[192] Therefore, a potential energy scan of this dihedral in ground state (orange) and excited state (black) has been obtained (see Figure 9.1.2). In order to gain the contribution of the bond turning potential to the potential energy of the dihedral angle, a scan (every 10°) of the dihedral by using MM optimization without a potential on the dihedral has been done with the GROMACS program package.[202] The thereby determined single point energy of the bond turning is subtracted from the CC2 scan and the curve thus obtained is fitted with the function: $F = a_0 + a_1 \cos(\frac{2x\pi}{180}) + a_2 \cos(\frac{x\pi}{180})$, while a_1 and a_2 (divided by 4) are the fitted potential force constants for the energy barrier of the dihedral. The CC2 scans and the scans received by the parametrized dihedral and with the GROMACS program package are shown in Fig 9.1.2. All parameters of the Flugi-2 molecule can be foudn in the appendix Table B

MD Simulation and Excitation Energies

The parametrized molecules (ground and excited state) have been simulated in gas phase and in solvent (DMSO) with a duration of 10 ns and a time step of 2 fs using the GROMACS package version 5.0.4. The Flugi-2 molecule was embedded in a cubic solvent box with a side length of 5 nm. Afterwards, the whole box was subjected to MM equilibration. In this equilibration, the first step was a minimization of the potential energy of the system with the steepest descent algorithm of 1000 steps and the maximum force tolerance was set to 100 kcal/mol.

A NVT ensemble over 1 ns and with a position restrained (force constant 1000 kJ mol^{-1} nm^{-1}) chromophore at a temperature of 300 K have been performed as second equilibration step.

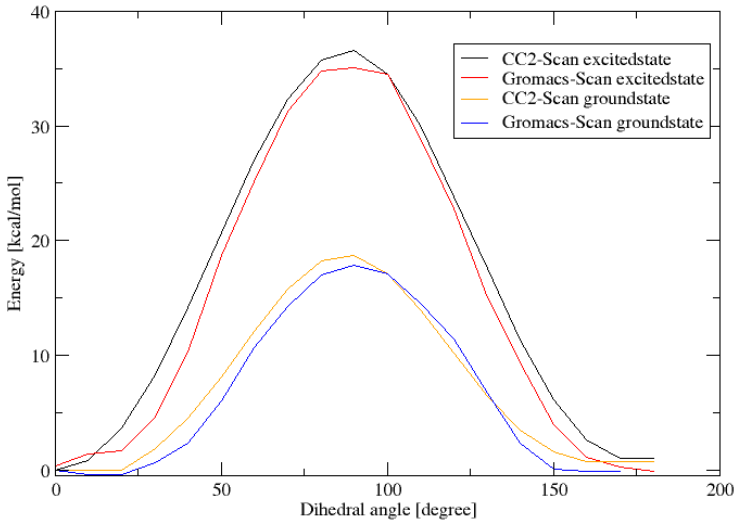


Figure 9.1.2: Scan of the dihedral angle N2-C8-C10-C14 with CC2 method in ground state (orange) and in excited state (black) and scan of MM-parametrized dihedral angle in ground state (blue) and excited state (red).

Initial velocities have been generated by the Maxwell-Boltzmann distribution at 300 K. The next equilibration step was in NPT ensemble over 10 ns and 1 bar pressure scaled by a Parinello-Rahman barostat (τ_p=2.0 ps, β=4.5 $\times$ 10^{-5}) and with restrained chromophore. The temperature was controlled at 300 K by the Nosé-Hoover thermostat. This procedure is a usual procedure for equilibrating the solvent. And for the non-bonded interaction, the group cut-off scheme has been chosen. After the equilibration the production run with a 20 ns length in a NPT ensemble at 300 K has been performed and the same settings as in the equilibration run have been applied, except of the position restrains on the chromophore. The chromophore has been released in the production run. The same settings have been used for the excited chromophore.

For comparison reason, the chromophore geometry has been optimized with the DFTB3 method and additionally, a QM/MM simulation in the ground state of the Flugi-2 molecule over 10 ns were performed, using the DFTB3 method[88] with 3ob parameters.[117,203] Furthermore, single point energies have been calculated with DFTB3, with a modified DFTB+ version,[204] using the DFTB3 method[88] with 3ob parameters[117,203] and the long-range-corrected functionals.[205]

Every 10 ps of the trajectory a snapshot was used to calculate the excitation energy with OM2/MRCI. The averaged excitation energies over the whole trajectory have been summarized in histograms, where the intensity of the bars has been determined based on frequency of the excitation energies times the oscillator strengths of the excitation. The

histograms are fitted to a gaussian $g(x) = a_0 \exp(a_2(a_1 - x)^2)$. The OM2/MRCI calculations have been performed with the MNDO program package[122] with 20 electrons in 20 orbitals, including single and double excitations.

9.2 Results

In this section the properties and structural parameters of the Flugi-2 molecule in the ground and excited state obtained by QM methods will be analyzed and discussed. Furthermore, the structural properties of the dynamical simulations, which have been obtained with the parametrized force field for Flugi-2 will be compared to the reference CC2-structure. The newly effects of Flugi-2 in DMSO solvent on the structure and the excitation energy will be discussed.

9.2.1 Properties of the Ground and Excited State Analyzed by QM Methods

The properties of the Flugi-2 molecule in the excited state are unknown, thus we performed various quantum mechanical calculations and analyzed the optimized CC2 structures to get an insight into the structural behavior of the molecule in the excited state. This knowledge is important to set up a reasonable parametrization of the molecule in the ground and excited state. Usually, conjugated π-systems like retinal invert their bond lengths in the excited state, that means the double bonds (DB) get longer and have a more single bond (SB) like character, while the single bonds get shorter and have more double bond character. In order to measure these bond lengths inversions of the conjugated double bonds in the excited state the bond length alternation (BLA) has been used. It is the difference between the averaged double bonds and the averaged single bonds of the conjugated π-systems. The BLA is directly correlated with the excitation energy, this means, the higher the BLA the further blue-shifted is the excitation energy.

The question arises, whether this bond length inversion occurs in the Flugi-2 as well? Therefore, the bond lengths of the optimized CC2 geometries of the excited state S_1 and ground state S_0 have been analyzed to find the structural changes between the two states.

It emerges, that there is no change of the bond lengths in the phenyl ring of the Flugi-2 molecule in the excited state, while in the imidazopyridine rings a change of the bond length (see Table 9.2.1 and Figure 9.2.1) has been found. Thus, the bonds with the largest change in the bond length between the ground and the excited state have to be identified (marked red in Table 9.2.1).

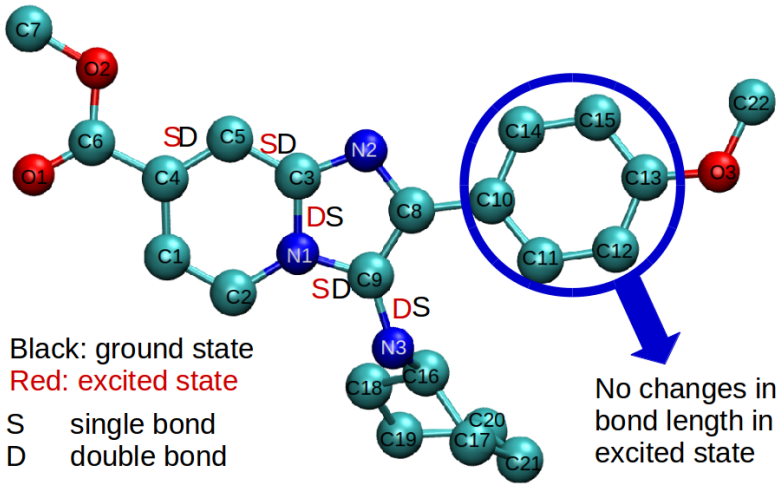


Figure 9.2.1: Bond length change between ground and excited state. The bond character is described by "S" single bond and "D" double bond. The bond lengths in ground and excited state are shown in Table 9.2.1.

The N1-C3 and C9-N3 single bonds show a decrease of the bond length of 0.04 Å. A standard C-N single bond in a conjugated π-system has a length of around 1.37 Å while a standard C=N double bond is approximately 1.33 Å long.[206] The bond distance of the C9-N3 bonds fit well with the standard bonds (see Table 9.2.1). Due to the large change of bond length inversion in the excited state, the C9-N3 bond can be identified as one of the bond inversions, the bond changes from a double to a single bond (D $\rightarrow$ S). However, the bond length of the N1-C3 bond is with a length of 1.42 Å in the ground state unusually long, compared to reference values of this bond type.[206] However, this bond length in the ground state has been confirmed by the crystal structure.[206] It has a similar length as a C-C bond in a conjugated system, while the value in the excited state (1.38 Å) matches with experimental measurements.[206] This could be due to the delocalization of the π-system in the imidazopyridine ring. Nevertheless, the value of the N1-C3 bond in ground state is in good agreement with the crystal structure of the 2-Phenylimidazo[1.2-a]pyridine[206] and DFT calculations.[207] It decreases in the excited state and can be identified as a single to double bond inversion.

The bonds C3-C5 and C5=C4 are 1.39 Å long in the ground state and fit well with the standard value of a C-C bond in a conjugated system (single C-C bond: 1.34-1.40 Å and double C-C bond: 1.42-1.46 Å[206]). Additionally, an increase of the bond length of about 0.03-0.04 Å has been found and the bond character switches from a single to a double bond. Similar effects have been observed for the C9-N1 bond. The ground state value (1.37 Å)

Table 9.2.1: Bond length and C8-C10 dihedral (bottom) of CC2 optimized geometries in ground and excited state and the change of the bond length or dihedral angle (third column). The bond character is described by "SB" single bond and "DB" double bond, while bonds, which change more than 0.02 Å are marked in red.

Bond length	S_0 [Å]		S_1 [Å]		Difference
C1=C2	1.37		1.39		0.02
C2-N1	1.37		1.38		0.01
N1-C3	1.42	SB	1.38	DB	0.04
C3-C5	1.39	DB	1.42	SB	0.03
C5=C4	1.39	DB	1.43	SB	0.04
C4-C1	1.42		1.40		0.02
C3=N2	1.34		1.36		0.02
N2-C8	1.37		1.36		0.01
C8=C9	1.41		1.42		0.01
C9-N1	1.37	DB	1.42	SB	0.05
C8-C10	1.46		1.44		0.02
C10-C11	1.41		1.41		0
C11=C12	1.39		1.39		0
C12-C13	1.40		1.40		0
C13=C15	1.40		1.40		0
C15-C14	1.39		1.39		0
C14=C10	1.40		1.41		0.01
C4-C6	1.48		1.46		0.02
C13-O3	1.36		1.36		0
C9-N3	1.39	SB	1.35	DB	0.04
Dihedral	S_0 [°]		S_1 [°]		
N2-C8-C10=C14	-18.70		-25.62		6.92
C9-N3-C16-C17	-59.09		-89.68		30.59

agrees well with experimentally determined bond lengths in an imidazole molecule.[206] The bond length increases about 0.05 Å in the excited state, this can be determined as an inversion of the bond character.

Thus, the bonds N1-C3, C9-N3, C3-C5, C5=C4 and C9-N1 are used to calculate the difference of the BLA with

$$BLA = \left(\frac{L(N1-C3) + L(C9-N3)}{2} - \frac{L(C3-C5) + L(C5-4) + L(C9-N1)}{3} \right). \quad (2.1)$$

The values for the optimized CC2-structure in ground and excited state are shown in Table 9.2.2. The value of the BLA decreases drastically in the excited state.

Table 9.2.2: Average single and double bond length of the CC2 optimized structures in ground and excited state and the calculated BLA. The average was computed by the bonds with a bond character inversion in the excited state (marked in red in Table 9.2.1 and illustrated in Figure 9.2.1)

	Ground state	Excited state
average DB [Å]	1.386	1.423
average SB [Å]	1.405	1.364
BLA	0.019	-0.059

Charges

The OM2/MRCI and SORCI calculations of the excited state show a large change of the dipole moment in the excited state. The total dipole moment in the ground state is about 2.1 D and in the excited state around 7.6 D. This indicates a change of the charge distribution within the Flugi-2 molecule.

The electron density and the charge distribution over the molecule have been calculated with the DFT-method in order to fit the atomic charges of the molecule. The fitted atomic charges in the ground state of the heavy atoms are shown in the appendix in Table 7. For a better comparison the sum of the charges in the different regions in the Flugi-2 molecule has been calculated (see Table 9.2.3). The major change between the ground state and excited state occurs at the cyclohexane. In the ground state the cyclohexane is almost uncharged, thus the interaction with an nonpolar solvent wont be strong. In the excited state the electron density is reduced at the cyclohexane and gets slightly positively charged. Therefore, a nonpolar solvent would interact with the transformed charge density at the cyclohexane in a different way than in the ground state. This gives a hint, why the Flugi-2 molecule is environmentally sensitive in the excited state and not in the ground state.

Table 9.2.3: Sum of point charges of different regions in the Flugi-2 molecule. The point charges are condensed on atoms derived from DFT calculations (details see subsubsection 9.1).

Part of Flugi-2	S_0	S_1	Difference
Imidazopyridine ring	-0.57	-0.51	-0.06
Benzene ring	-0.42	-0.40	-0.01
cyclohexane-ring without N3	0.18	1.54	-1.35
N3	-0.85	-1.23	0.38
C6-O1-O2-Me	-0.19	-0.16	-0.03
O3-C22	-0.33	-0.91	0.58

9.2.2 Molecular Dynamic Simulation in the Ground State

In the last subsection 9.2.1 the two states (ground and excited state) of the Flugi-2 molecule have been characterized and the transformations, which occur due to the transition from the ground to the excited state have been described and understood. As mentioned at the beginning of this chapter, a lot more could occur in the chromophore, when considering the dynamics of the system. To understand the complete absorption and emission process an important effect, the dynamical behavior of the molecule, is still missing. The strategy is to include these dynamical behavior of the molecule by classical MD simulations. Thereby, a set of structures or geometries, called trajectory, which reproduce the fluctuation and dynamic effect in the ground state will be obtained. This will give an insight in the dynamical processes in the molecule. The excitation energy calculations of this trajectory include now the dynamic interactions of the Flugi-2 and their contribution to the excitation energy. Therefore, the Flugi-2 molecule has to be parametrized in order to propagate classical force field simulations. The parametrization has been done based on the CC2-structures (details see section 9.1). However, contrary to the QM methods, the classical description makes use of some approximations, or rather it applies parametrized values to specify bond lengths, angles, dihedral angles and atomic charges of the molecule (more detailed description see chapter 5). These structural parameters have a direct influence on the excitation energy of the molecule. Therefore, the parametrization has to be very accurate and has to be intensively tested and validated by QM methods. To achieve this, the BLA, the dihedral angles N2-C8-C10-C14 and C9-N3-C16-C17 and the excitation energy calculated with OM2/MRCI method have been chosen for a comparison with QM optimized structures (see Table 9.2.4).

The classical force field minimized structure (MM-minimized) underestimates the BLA compared to the CC2 reference structure. In the MM-minimized structure, the averaged single bonds are much shorter than in the QM minimized structures (CC2 and DFTB). The

analysis of the single bond lengths in the MM-minimized structure shows, that the force field underestimates the N1-C3 single bond length (1.39 Å) (details see Table 8 in the appendix) and this results in the underestimation of the BLA. Several bond length parameters (b_0) for this bond length have been tested, but the reference bond length of the CC2-structure has never been reached. This wrong parametrized bond length is responsible for the shift of the excitation energy about 0.12 eV to lower energy compared to the excitation energy of the CC2-structure.

Table 9.2.4: Comparison of structural properties of the minimized structure by MM (first column) with CC2-reference-structure (last column) and DFTB-structure in the ground state.

	MM minimized	DFTB(3ob)	CC2
average DB [Å]	1.387	1.391	1.386
average SB [Å]	1.390	1.418	1.405
BLA	0.003	0.027	0.019
N2-C8-C10-C14 [°]	-29.81	-15.74	-18.70
C9-N3-C16-C17 [°]	-86.68	-65.95	-59.09
OM2/MRCI [eV]	3.45	3.54	3.57

In the next step, a 10 ns MD simulation of the Flugi-2 molecule in gas phase has been performed and analyzed. Table 9.2.5 shows the averaged BLA of the MD simulation in the ground state without a solvent. As well as for the MM-minimized structure (Table 9.2.4), the averaged BLA of the MM-MD trajectory is too small compared to the CC2-structure and the averaged excitation energy is red shifted, too. However, comparing the excitation energy of the MM-minimized structure and the MM-MD trajectory, the averaged excitation energy of the MM-MD is red shifted about 0.07 eV. In other words, it this shifted about 0.2 eV compared to the CC2-structure. This red shift cannot only be explained by the BLA. Because considering the fact that the BLA and the excitation energy show a linear relation, and the BLA between CC2-structure and MM-minimized structure change about 0.016, while the excitation energy changes about 0.12 eV. Thus, the change of the BLA (0.0008) is too small for this red shift of 0.07 eV. The excitation energy shift is probably due to a dynamic effect. Even in the DFTB-MD simulation a red shift of 0.06 eV compared to the DFTB minimized structure has been observed. This shift does not result from a parametrization, but results from dynamical structural variations within the trajectory.

Plotting the distribution of the dihedral angles of N2-C8-C10-C14 and C9-N3-C16-C17 of the MM-MD and the DFTB-MD in Figure 9.2.2, a flexible Flugi-2 molecule has been observed in both simulations. This might result in an excitation energy shift, if there is a correlation between dihedral angles and the excitation energy. The excitation energies have been plotted versus the dihedral angle (see Figure 9.2.3) and, with a k-means cluster

Table 9.2.5: Comparison of averaged BLA and the averaged excitation energy with OM2/MRCI of the MM-MD over 10 ns (first column), DFTB-MD over 1 ns (second column) and minimized CC2 structure (last column) in the ground state

	MM-MD	DFTB(3ob)-MD	CC2
BLA	0.002	0.026	0.019
OM2/MRCI [eV]	3.38	3.48	3.57

analysis, the cluster centroids have been computed (black symbols in Figure 9.2.3). A minor correlation between the dihedral angle and the excitation energy has been observed. Thus, the excitation energy is not correlated to the dihedral C9-N3-C16-C17 (9.2.3a) and N2-C8-C10-C14 (9.2.3b). Since the MD simulations in gas phase and DMSO sample different dihedral angle configurations, a small shift in the averaged excitation energy can be observed.

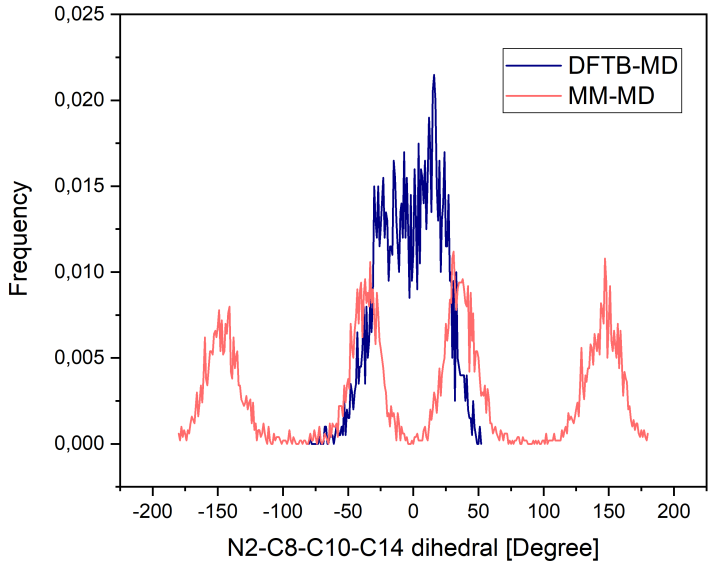
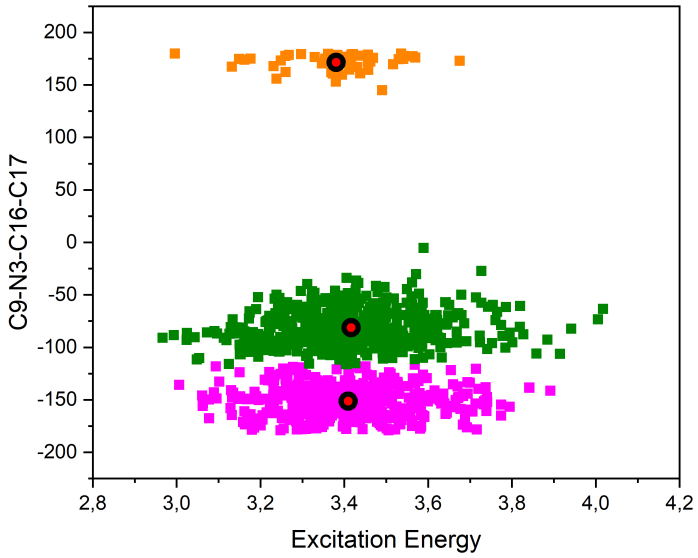
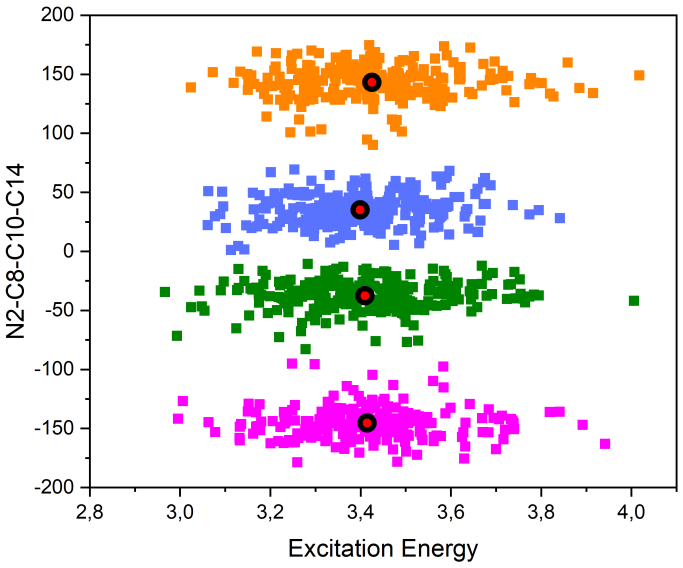


Figure 9.2.2: Distribution of the N2-C8-C10-C14 dihedral angle over 10 ns simulation time with MM-MD (red) and over 1 ns simulation time with DFTB-MD (blue).

Another correlation can be between the BLA and the excitation energy. Further analysis of the BLA of the MM-MD show, that there is a correlation between several bonds and the excitation energy. In Figure 9.2.4 the linear fits of the bond length distributions over the 10 ns MM-MD simulation have been shown. The bond length has been plotted against the excitation energy and the data have been fitted by a linear approach. In Figure 9.2.4 the bond lengths, with a gradient of the linear fit of 0 or within the standard deviation, have not



(a)



(b)

Figure 9.2.3: Plot of excitation energy vs dihedral angles of a) C9-N3-C16-C17 and b) N2-C8-C10-C14 of the 10 ns MM-MD in gas phase. There is no correlation distinguishable.

been shown, because in this bond lengths no correlation between the bond length and the excitation energy can be found. The increase of the bond lengths C9-N3, C6-C4, N2-C8, C1-C4, C8-C10 and C3-N1 result in a blue shift of the excitation energy, while the increase of the C8-C9, C4-C5, C3-N2, C9-N1 and C2-C1 bond lengths result in a red shift of the excitation energy. Since there are small variations in the averaged bond lengths during the 10 ns simulation, these bonds might be also responsible for the excitation energy shift of the MM-MD compared to the MM-minimized structure.

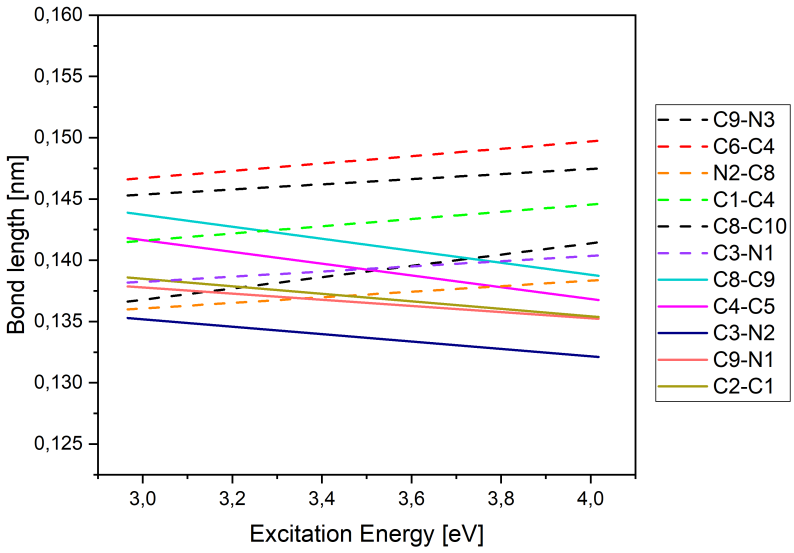


Figure 9.2.4: Linear fits of the bond length distribution over 10 ns simulation time. Here are shown only the bond lengths with a correlation to the excitation energy, which means the gradient has to be larger than 0.0005 nm and larger than the standard deviation of the fit.

9.2.3 Molecular Dynamic Simulation in the Excited State

For the excited state, the parametrization needs to be validate as well. The excitation energy of the MM structure as well as the excitation energy of the DFT(ωB97X) and DFTB are around 0.1 eV blue shifted compared to the excitation energies of the CC2 structure. What are the structural rearrangements inducing this shift? In the ground state a minor correlation of the dihedral angle of the cyclohexane and the phenyl and the excitation energy has been observed. In Table 9.2.6, the structural properties and the excitation energy of the structures optimized in the excited state by different methods have been shown. It emerges, that the excitation energy is independent of the dihedral angles. Here, the dihedral angles are not the root of the blue shift.

Furthermore, in the ground state a correlation of several bond lengths with the excitation energy has been found in the MM-MD simulation. This has not been found in the excited

state. The same computed linear fits for the bond lengths in the excited state of the MM-MD, which were done for the bond lengths in the ground state (Figure 9.2.4) give only a correlation of the C4-C5 and C9-N1 bond with the excitation energy. This means, that the excitation energy in the excited state seems to be mostly independent of the bond lengths or is compensated by other excitation energy correlated structural properties, for example dihedral angles.

Table 9.2.6: Comparison of structural properties of the MM-minimized structure and the averaged MM-MD structure with CC2-structure. DFTB-structure and a DFT(ωB97X)-structure for the first excited state.

	MM minimized	MM-MD	DFT-ωB97X	DFTB (3ob)	CC2
average DB [Å]	1.428	1.428	1.406	1.430	1.423
average SB [Å]	1.361	1.356	1.367	1.379	1.364
BLA	0.067	0.072	0.039	0.051	0.060
N2-C8-C10-C14 [°]	-29.81	-24.93	-27.10	-23.70	-25.62
C9-N3-C16-C17 [°]	-90.44	-39.61	-86.52	-93.38	-89.68
OM2/MRCI [eV]	3.01	2.96	3.01	3.04	2.88

The averaged excitation energy of the MM-MD over 10 ns simulation time is red shifted about 0.05 eV compared to the excitation energy of the MM-minimized structure (3.01 eV) (see Table 9.2.6). This has been already observed for the ground state. Since there is no correlation of the bond lengths and the dihedral angle to the excitation energy, it is not clear what the reasons for this shift are. Nevertheless, the shift lies in the error range of our method OM2/MRCI.

Another parameter is the N2-C8-C10-C14 dihedral angle, this dihedral angle has been parametrized for the Flugi-2 molecule. The potential curves have been shown in Figure 9.1.2. The rotation barrier of the dihedral angle in the excited state is much higher than the barrier in the ground state. Thus, in the ground state the rotation around N2-C8-C10-C14 dihedral is more flexible and visible in the distribution of the dihedral over the simulation time. Figure 9.2.5 compares the occurence of the of the N2-C8-C10-C14 dihedral angles in the ground (black) and in the excited state (red) and confirms this assumption.

In the last sections, the structural reason for the color shift between the ground and excited state has been found. Furthermore, the relation between excitation energy and structural parameters of MM-MDs of ground and excited state have been analyzed. But the absolute values of the excitation energy have not been reproduced by MM methods. The reason for this inaccuracy is that generating a force field which describes the correct bond lengths

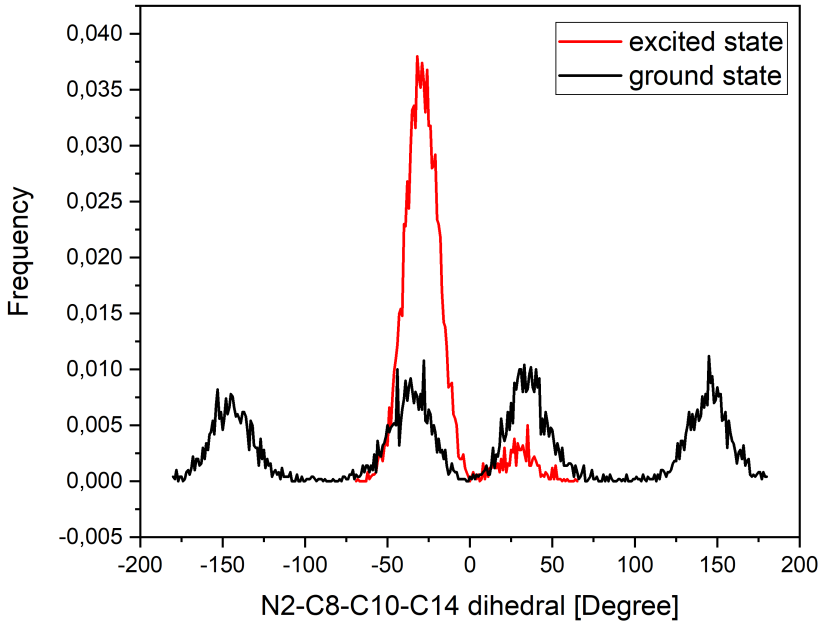


Figure 9.2.5: Distribution of N2-C8-C10-C14 dihedral in the ground (black) and in the excited state (red) over 20 ns simulation time.

and dihedral angles in conjugated π-systems is a challenging task. Too many effects need to be considered and are only approximated in the MM application. Even QM methods like DFT and DFTB fail when calculating the right bond lengths and angles (Table 9.2.6) of the molecule. However, the exact absolute absorption and fluorescence energy is not the most important topic of the investigation. Much more important is the relation between structure and excitation energy. For this approach the MM methods can be a useful tool, as illustrated by finding the correlation between bond length, dihedral angle and excitation energy (Figure 9.2.4).

Nevertheless, beside the absolute value of the excitation energy, the Stokes shift is a interesting property when investigating fluorescence processes. Table 9.2.7 shows a summary of the absorption and the fluorescence energies computed with different methods in gas phase.

If only difference in energies, the Stokes shift, is compared, it can be seen that the CC2-structure agrees well with the experimental values in decane, while DFT structure overestimates the Stokes shift about 0.18 eV and the structure of the semi-empirical method DFTB underestimates the Stokes shift about 0.15 eV. Even in the gas phase we get a high variation of the Stokes shift between different QM methods. These energy variations are connected to variations in the structure of the molecule and this has been analyzed in the last sections.

The MM method underestimates the Stokes shift about 0.24 eV. This is in bad agreement with the CC2-structure in the ground state. The reason for this underestimation of the Stokes shift of the MM-structures is because the parametrization failed to describe the right BLA in the imidazopyridine ring in the ground state. This results in a shift of about 0.2 eV of the absorption energy compared to the CC2 value.

The shift in the excited state is only about 0.08 eV, thus, the MM-MD for the excited state is usable for further investigations, while the ground state MM-MD is not.

Table 9.2.7: Comparison of the OM2/MRCI excitation energies of the ground state (absorption) and excited state (fluorescence) structures received by different methods as well as the consequently calculated Stokes shift (last column). The Stokes shift is not reproduced by the MM structures. In the last row are shown the experimental values of the Absorption and Fluorescence in decane.

Structure	Absorption [eV]	Fluorescence [eV]	Stoke Shift [eV]
MM-MD	3.37	2.96	0.41
MM-minimized	3.45	3.01	0.44
DFTB 3ob	3.54	3.04	0.50
DFT-ωB97X	3.84	3.01	0.83
CC2-structure	3.57	2.88	0.69
Experiment (in decane)	3.26	2.61	0.65

9.2.4 Influence of Solvent DMSO on the Excited State Dynamics

As mentioned at the beginning of the chapter the Flugi-2 molecule has an interesting feature, the fluorescence energy and accordingly the Stokes shift is dependent on the environment of the chromophore. The interaction of the Flugi-2 with the solvent in the excited state leads to a red shift of the fluorescence wavelength. For example in decane the fluorescence wavelength is at 380 nm, which is equal to 2.61 eV, and in DMSO the fluorescence wavelength is shifted to 540 nm or 2.30 eV, this gives a shift of about 0.31 eV and is the highest experimentally observed solvent dependent shift. In contrast to most chromophores the environmental sensitivity is only observed for the fluorescence process, while in the ground state the solvent does not affect the absorption wavelength (except DMSO). In more detail, some structural and electrostatic rearrangements must occur within the chromophore in the excited state, so that the solvent or the environment influences the molecule only in the excited state, whereas the chromophore in ground state is mostly unaffected by the solvent.

A first hint had been given by the calculated charge distribution in the excited state seen in Table 9.2.3. The charge distribution in the excited state has mainly changed in the cyclohexane ring and at the N3 atom. The charges between these two molecule parts in

the excited state are more separated than in the ground state. Thus, the interaction in the cyclohexane ring and the N3-atom with a polar solvent molecule has to be increased in the excited state.

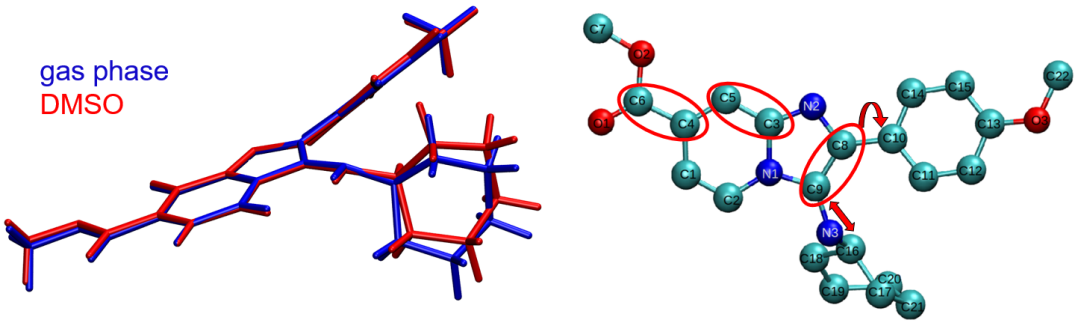


Figure 9.2.6: Flugi-2 structure optimized with DFT (ωB97X) in gas phase (blue) and by switch on implicit solvent model COSMO for DMSO (red). The main changes in the structure happen on the bond length of C3-C5, C8=9, C4-C6 bond and on the dihedral C9-N3-C16-C17 and N2-C8-C10=C14 (see Table 9.2.8).

In addition, the gas phase structure in the excited state gives no hint why the polar solvent should interact differently with the Flugi-2 molecule in the excited state than in the ground state. Possibly, the polar solvent has an impact on the structure in the excited state. Unfortunately, it is not easy to optimize a structure in the excited state by QM methods with explicit solvent molecules, because only implicit solvent models are available and there the solvent is treated as an indirect matter on the surface of the chromophore. The solvent molecules act like a constant background and the chromophore 'sees' the solvent only through electrostatic interaction. Another option to consider the solvent environment is to include the solvent molecules in the QM calculation as point charges, but here the charges are fixed on their locations during the QM calculation and to get the locations a additional MM calculation is necessary to determine the locations of the charges in advance.

Nevertheless, it might be that in the structure optimized with implicit solvent models already differences of the structure occur. The Flugi-2 has been optimized with DFT and a range-separated ωB97X functional, using the COSMO model with DMSO ($\epsilon = 47.2$ and refraction parameter$= 1.479$). A comparison of the gas phase and DMSO structure has been shown in Figure 9.2.6 and Table 9.2.8.

The main differences are in the different dihedral angles of the cyclohexane ring and changes of some bond lengths. This structural rearrangements lead to a small red shift of the excitation energy as seen in Table 9.2.9. Here, the excitation energies of the optimized structures in gas phase and DMSO are compared with different methods. The strength of the red shift is 0.07 eV for the OM2/MRCI method and the SORCI method in gas phase, while with DFT/ωB97X the shift due to the structural changes is only 0.04 eV. This shows, that not only the structural changes are responsible for the red-shift.

Table 9.2.8: Bond lengths and C8-C10, N3-C16 dihedral angles (bottom) of DFT/ωWb97X optimized geometries in excited state in gas phase, with COSMO (DMSO) and with one DMSO molecule which is hydrogen bonded to the N3-H atoms of the Flugi-2 molecule.

Bond length	gas phase [Å]	COSMO(DMSO) [Å]	H-bonded DMSO [Å]
C1=C2	1.38	1.37	1,39
C2-N1	1.39	1.40	1,39
N1-C3	1.38	1.39	1,38
C3-C5	1.40	1.38	1,40
C5=C4	1.43	1.44	1,43
C4-C1	1.40	1.41	1,40
C8=C9	1.42	1.44	1,42
C9-N1	1.39	1.38	1,40
C4-C6	1.47	1.45	1,47
C9-N3	1.35	1.34	1,34

Dihedral	gas phase [Degree]	COSMO(DMSO) [Degree]	H-bonded DMSO [Degree]
C9-N3-C16-C17	-100.58	-86.52	-50,68
N2-C8-C10-C14	-27.1	-32.54	-36,76
C9-C8-C10-C11	-26.94	-33.02	-36,85

The electrostatic influence of the implicit solvent is shown in the columns, when comparing the excitation energies of the same structure in gas phase and with the COSMO model in DMSO. The red shift of the excitation energy due to the implicit solvent is around 0.15 eV (3.35 eV-3.20 eV) for DFT/ωB97X and less than 0.04 eV for OM2/MRCI and SORCI. Already with different QM methods a varying shift is observed. It could be that DFT/ωB97X overestimates this shift, while OM2/MRCI and SORCI underestimates it. Clarifying this needs additional excitation energy calculation with higher level QM methods like CC2. When combining the structural rearrangements (columns) and the electrostatic interactions (rows) in the excitation energy calculations a red shift of 0.21 eV with DFT/ωB97X and of 0.2 eV with OM2/MRCI can be achieved. This is much closer to the experimental red-shift about 0.31 eV. So the structural changes and the electrostatic effects are important for the red shift, but probably not all effects are yet considered by the QM methods.

Table 9.2.9: Here, the excitation energies of the excited state (fluorescence) geometries are shown. They geometries are optimized in gas phase with CC2 and DFT/ωB97X, in DMSO by using the implicit solvent model COSMO (ωB97X DMSO). Furthermore a Flugi-2 with one DMSO molecule, which is hydrogen bonded to the N3-H atoms structure has been optimized with DFT/ωB97X. In the columns the excitation energies of the different optimized structures are shown and in the rows are list the excitation energies gained by different methods.

Excitation method \ Optimization Method	CC2 gas phase	ωB97X gas phase	ωB97X DMSO	ωB97X H-bonded DMSO
CC2 gas phase	2.56	-	-	-
SORCI gas phase	2.37	2.50	2.43	-
SORCI DMSO	2.35	2.50	2.44	-
DFT/ωB97X gas phase	3.22	3.35	3.31	3.17
DFT/ωB97X DMSO	3.05	3.20	3.14	3.07
OM2/MRCI gas phase	2.88	3.01	2.94	2.88
OM2/MRCI DMSO	2.87	2.97	2.81	2.87

For further analysis of the red-shift of the fluorescence energy, the chromophore has been added in a box of DMSO molecules and a 20 ns simulation has been performed. For the Flugi-2 molecule the gas phase parametrization of the excited state has been used. Every 10 ps of the simulation time, a snapshot has been used for a single point excitation energy calculation with OM2/MRCI. In order to generate a spectrum, the 2000 fluorescence energies have been combined in a histogram, where the intensity is defined by the counts of the energy in a certain segment, times the oscillator strength resulting from the OM2/MRCI fluorescence energy calculation. Figure 9.2.7 shows, that the fluorescence energy maximum

is shifted about 0.21 eV due to the solvent DMSO. Still, the experimental shift is not reached. Comparing the MM-minimized structures only a 0.06 eV red shift (Table 9.2.10) has been achieved. Why has a shift in the MM-MD been observed, while in the MM-minimized structure no shift has been observed? The BLA of the MM-minimized and the MM-MD structures do not explain the increased red shift in the MM-MD, contrariwise the BLA changes in the MM-minimized structures and in the MM-MD trajectory are almost equal. Thus, another explanation for the large red-shift in the MM-MD should exist.

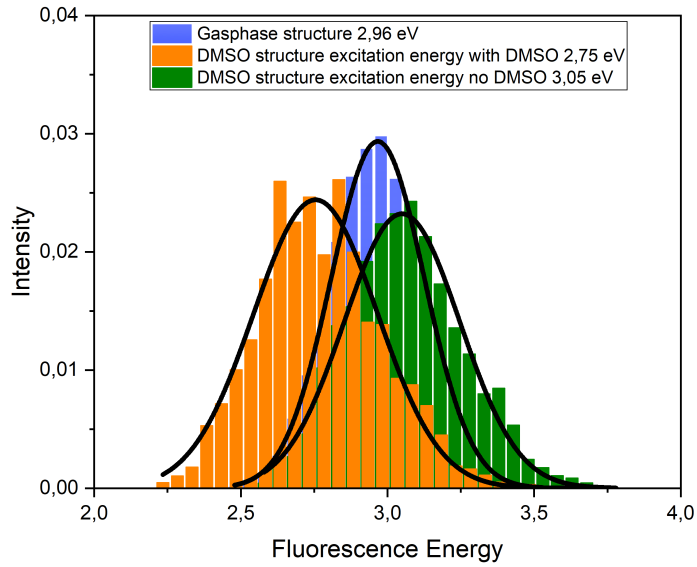
Table 9.2.10: The excitation energies of the excited state (fluorescence) geometry are shown. Also, the experimental values are listed. The second row shows the excitation energy from a structure minimized by MM force field. In this structure the hydrogen bond as described in Figure 9.2.8 is not present. The excitation energies are calculated by OM2/MRCI and considering DMSO as point charges.

fluorescence energy and BLA	gas phase	DMSO	Difference
MM-MD [eV]	2.96	2.74	0.22
BLA MM-MD [Å]	0.072	0.076	0.004
MM minimized [eV]	3.01	3.07	0.06
BLA MM minimized [Å]	0.067	0.074	0.007
Experiment [eV]	2.61 (in decane)	2.30 (in DMSO)	0.31

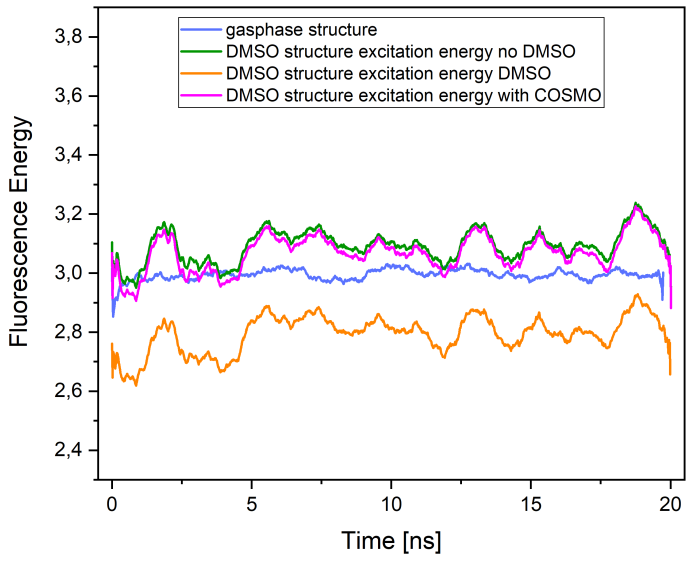
The investigation of the excitation energy over time in 9.2.7b shows more fluctuations in the DMSO trajectory (orange) than in the gas phase trajectory (blue). Moreover, the excitation energy plot of the DMSO trajectory is about 0.2 eV red shifted. Here, the DMSO molecules are included in the calculation of the excitation energy as point charges. If only the Flugi-2 structure in the DMSO had been used and the electrostatic effect of the DMSO had been neglected (removing the point charges), the fluctuations would still be present, but the averaged excitation energy would be slightly blue shifted (green curve in 9.2.7a) compared to the gas phase trajectory.

To find out how important the explicit description of the electrostatic interaction is, the electrostatic influence is included again in the OM2/MRCI calculation, though not as explicit point charges, but as implicit solvent model COSMO (pink curve). The excitation energy is not back-shifted to the initial value of 2.74 eV, only a minor red-shift is noticeable in the graph. Thus, the point charges are very important in the calculations and can not be replaced by a solvent model. Furthermore, the structural fluctuations have an impact on the excitation energy, too. However, this impact results in a slightly blue shift, therefore neither of this interactions can be neglected.

A more detailed look at the trajectory shows a hydrogen bond between the oxygen atom of a DMSO molecule and the N3-H atoms of the Flugi-2 molecule. This hydrogen bond is stable over the 20 ns simulation time and influences the C8-C10 and N3-C16 dihedral



(a)



(b)

Figure 9.2.7: a) Histogram of OM2/MRCI excitation energies and b) excitation energy plotted over time respectively of MM-MD-geometries in excited state of the gas phase MM-MD (blue) and of the MM-MD in DMSO with the electrostatic interaction of the DMSO by point charges (orange) and without (green).

angles (Figure 9.2.8) very strongly.

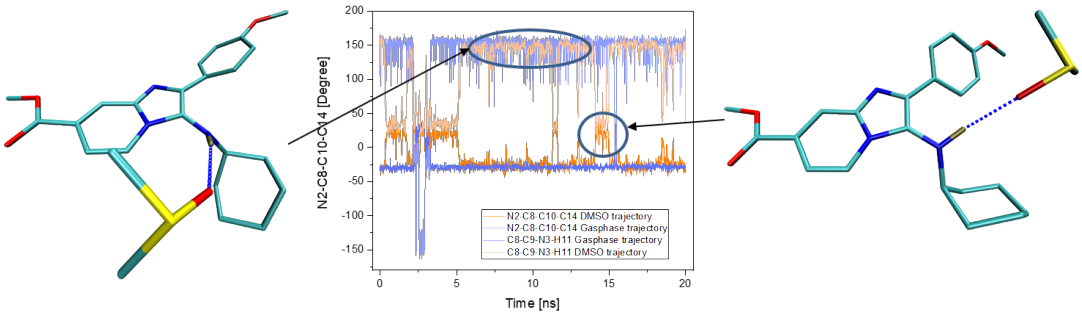
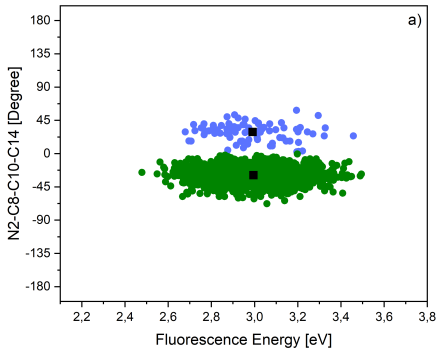


Figure 9.2.8: C8-C10 and N3-C16 dihedral angles over 20 ns simulation time.

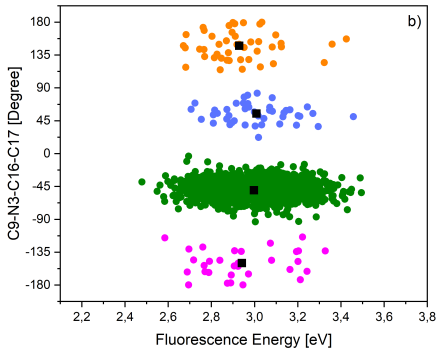
The plot of the excitation energy against the dihedral angles and a following k-means cluster analysis shows a correlation between the dihedral angles and the excitation energy in DMSO (see 9.2.9c and 9.2.9d). The centroid points for each cluster (black points in Figure 9.2.9) mark the mean value of the excitation energy for each cluster. Comparing the centroid points for the DMSO trajectory with the gas phase trajectory, the dihedral sampling area of the C9-N3-C16-C17 dihedral angle is shifted to lower dihedral angles, while the region between 0-135° is not sampled at all in the DMSO trajectory (9.2.9d). This could result from the hydrogen bond of the N3-H atoms to a DMSO molecule.

Moreover, a red shift of the excitation energy is observed for the individual clusters of the C9-N3-C16-C17 and N2-C8-C10-C14 dihedral angles in the DMSO trajectory (see Figure 9.2.9). But comparing the centroid points of the clusters with the same dihedral angle between the gas phase (9.2.9b) and the DMSO trajectory (9.2.9d) shows that the red shift can not be exclusively attributed to the dihedral angles. The centroid points of the same clusters in gas phase and DMSO (for example the cluster around $-45°$ in 9.2.9b and 9.2.9d) do not correspond to the same excitation energy. This means the excitation energies with the same dihedral angle are different in gas phase and DMSO. However, the angle distribution in the DMSO trajectory is different to the gas phase trajectory and could be a reason for the red shift of the excitation energy in DMSO. The distribution of the dihedral angles in gas phase trajectory is less spread out. The dihedral angles in gas phase do not fluctuate as strongly as in the DMSO trajectory, because of the lacking interaction with DMSO molecules. Furthermore, the hydrogen bond to the DMSO enables the Flugi-2 to sample more dihedral angles than in gas phase.

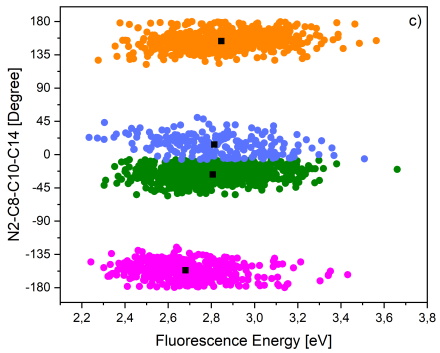
Another hint that the hydrogen bond is one of the main factors for the red shift of the excitation is the small red-shift of the excitation energy for the MM-minimized structure, because in this structure the hydrogen bond is not present and the red shift is almost vanished (see Table 9.2.10).



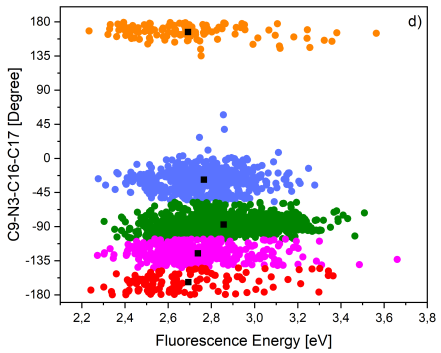
(a) Cluster analysis N2-C8-C10-C14 dihedral gas phase.



(b) Cluster analysis C9-N3-C16-C17 dihedral gas phase.



(c) Cluster analysis N2-C8-C10-C14 dihedral in DMSO.



(d) Cluster analysis C9-N3-C16-C17 dihedral in DMSO.

Figure 9.2.9: Dihedral angle of N2-C8-C10-C14 (a) and c) and C9-N3-C16-C17 (b) and d) plotted against excitation energy in gas phase (top) and DMSO (bottom). A k-means cluster analysis was applied to determine the centroid points (black points) for each cluster.

Additionally, the influence of the DMSO on the atomic charges of the Flugi-2 molecule in the excited state has been tested. To this end, the charges have been calculated with DFT using an implicit solvent model (COSMO). The sum of the charges does not change much when considering an implicit solvent model in the calculation (see appendix Table 7). The charges have been included in the parametrization and a 10 ns MM-MD simulation has been performed (see Figure 9.2.10). The fluorescence spectra of the trajectory with the COSMO-charges and the gas phase trajectory have been compared, the spectra show only a negligible difference. Therefore, the influence of the solvent on the atomic charges has been neglected in the parametrization.

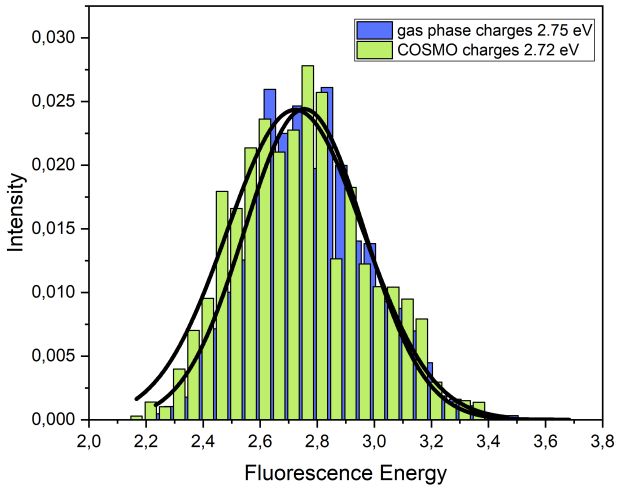


Figure 9.2.10: Histogramm of OM2 calculations of MM-MD-geometries in excited state/fluorescence in solvent (DMSO) using parametrization with gas phase charges (blue) and COSMO charges (green).

The most important interaction to explain the red shift in the excited state seems to be the hydrogen bond between DMSO and N3-H atoms of the Flugi-2 and the therefore changed C9-N3-C16-C17 dihedral angle. Thus, the Flugi-2 molecule and the single hydrogen bonded DMSO have been optimized with DFT/ωB97X and the structural changes (Table 9.2.8) and the excitation energies (Table 9.2.9) have been compared. The excitation energy shift from the gas phase structure to the structure with one DMSO is already 0.13 eV and is much higher than the red shift reached using an implicit solvent model. Still, the experimental shift is not reproduced, however the interaction of the surrounding DMSO molecules is not considered and the dynamic effects are neglected. Nevertheless, the explicit description of the interaction between the Flugi-2 and the DMSO is crucial to understand the red-shift of the fluorescence energy and cannot be described by an averaged polar surface interaction like in COSMO.

The structural rearrangements due to the hydrogen bond are very small. Except the N3-C16 dihedral angle, it changes compared to the gas phase structure around 50°, though the COSMO structure overestimates this turn about 30°. The charges of the hydrogen bonded structure have not been determined, yet. It is possible, that the explicitly described DMSO molecule influences the charge distribution of the whole Flugi-2 molecule much more than an implicit solvent model would influence it.

In Figure 9.2.10 only the atomic charges, received by a COSMO calculation have been tested and therefore the charge distribution could be wrong, since the COSMO model neglects some effects. The hydrogen bonded DMSO structure does not feel the electrostatic interaction of the DMSO solvent on the overall surface of the molecule, but still this single DMSO molecule has an high effect on the excitation energy. In the next step more DMSO molecules should be included to see how they influence the structure, the charges and the energy of the Flugi-2. To test this, the frozen-embedding calculations are useful, since here, the charges of the Flugi-2 molecule can be calculated, while the solvent is described explicitly.

9.2.5 Concluding Remarks

Concluding this chapter, the structural dependencies of the Flugi-2 molecule on the excitation energy in the ground state and in the excited state have been elucidated. Classical MD simulations combined with QM calculations have been used. It can be concluded, that molecular mechanics are a useful tool to sample geometries in order to get an insight in the structure-excitation energy relationships. Especially it was able to get an insight in the behavior of the Flugi-2 in the excited state due to the parametrization. For the Flugi-2 molecule a correlation between the dihedral angle of the cyclohexane and the excitation energy has been found. But nevertheless, it is difficult to sample the exact minima of the potential surfaces of such π-conjugated ring systems with force field methods. Even the lower level QM methods struggle describing the right structures of such systems. Thus, one has to be careful and validations and tests of the parametrization of MM force fields are necessary.

Moreover, the importance of explicit solvent interactions has been determined. The influence of the explicit solvent molecules should not be neglected when describing the connection between excitation energy and structure. Especially for the Flugi-2, the hydrogen bond pattern is important to get the right red-shift. Here, the molecular dynamics are very useful to investigate these structural patterns between solvent and chromophore. Furthermore, the implicit solvent models fail to describe the right interaction between solvent and chromophore.

10 | Conclusion and Outlook

The research of biological processes requires system independent, nonlethal, fast and microscopically visible tools. These requirements can be provided by systems interacting with light, like optogenetic methods or fluorescent markers. However, the investigation of the interaction between matter and light, opens new dimensions, at which the insight and understanding of these processes pave the way for new developments of chromophores and applications of photophysical and photochemical processes in biology and technology. In this thesis, these interactions have been investigated on absorption, fluorescence and isomerization processes in different environmental systems.

The connection between absorption wave length and spectral shape in channelrhodopsin-2 was investigated in chapter 6. It was shown, that the dynamic behavior of the binding pocket influences the spectral shape. Structural diversity found around the environmental sensitive chromophore leads to the broadening and the sub-peak character of the absorption spectrum in channelrhodopsin-2. The dynamic behavior almost vanishes at low temperatures and the structural diversity is not present anymore. This shows the strong temperature dependency of the experimental results. The analysis of the crystal structure of channelrhodopsin-2 depicts, that the presence and the amount of water molecules in proteins are important for the structural diversity and flexibility in proteins. The larger amount of water molecules leads to a more dynamic active site. The detection of side chain conformations with minimum energy has been achieved by MD simulations of several hundred nanoseconds and are verified by metadynamic free energy calculations. Since, the investigation is based on a homology model, possible rearrangements can not necessarily be interpreted as function mechanism rearrangements. However, occuring rearrangements have to be verified by free energy calculations. This is also the case for the selection of the QM-zone between protein and chromophore, since the interface between MM and QM can lead to artifacts and to incorrect descriptions of the structure.

In chapter 7 a homology model for HKR1 has been built and successfully validated by comparing the absorption with the reference bacteriorhodopsin. The prediction of the binding pocket and the retinal configuration leads to the characteristic shift of HKR1 absorption spectrum. The experimental blue shifted absorption of HKR1 with respect to bacteriorhodopsin was found to result from the non-planar retinal conformation, and not from

the amino acid composition in the binding pocket. After the calculation of the photochemical path, a reliable 13-*cis*, 15-*syn* structure has been achieved for the further investigation of the unique second stable state in the photocycle.

The semi-empirical methods like OM2/MRCI are computationally fast and thus a dynamic simulation of the isomerization process in rhodopsins is possible. Therefore, OM2/MRCI has been benchmarked for the prediction of the excited state life time and the isomerization mechanism on a reduced retinal model. OM2/MRCI excited state dynamics provide the correct prediction of the isomerization mechanism, but fail in giving the correct excited state life time. However, previous studies by Dokukina *et al.*[154] promise more reliable result for longer retinal models like PSB4 or PSB5.

In studying the fluorescence properties of the Flugi-2 molecule with the MM method, it has been found to that classical mechanics are a useful tool to investigate the structural interaction between chromophore and environment. The Classical method can provide geometries in the excited state, while semi-empirical methods are suitable to evaluate the electronic structure in order to get a meaningful picture of the interaction occuring in the excited state between chromophore and solvent. The simulations show, that the explicit solvent should not be neglected in the dynamics, because interactions between solvent molecules and chromophore are crucial for the correct description of the structure in the excited state and thus for the fluorescence wave length. However, force field methods are not able to include the electrostatic interactions, which leads to an underestimation of the solvent shift. Thus, future methods should include a quantum mechanical description of the chromophore and provide dynamic geometries in the nanosecond timescale in the excited state. Simultaneously, a classical description of the surrounding molecules or the protein is necessary to include the influence of the environment. The investigation of fluorescence properties and characteristics is a relatively new field and provides a great opportunities to develop new methodologies.

Bibliography

[1] P. Kln and J. Wirz, *Photochemistry of Organic Compounds*. Chichester, UK: John Wiley & Sons, Ltd, feb 2009.

[2] S. E. Braslavsky, "Glossary of terms used in photochemistry, 3rd edition (IUPAC Recommendations 2006)," *Pure and Applied Chemistry*, vol. 79, pp. 293–465, jan 2007.

[3] D. Wöhrle, M. W. Tausch, W.-D. Stohrer, and John Wiley & Sons., *Photochemie : Konzepte, Methoden, Experimente*. Wiley-VCH, 1998.

[4] W. W. Parson, *Modern Optical Spectroscopy.* Springer-Verlag Berlin AN, 2016.

[5] A. Jabłoński, "Über den Mechanismus der Photolumineszenz von Farbstoffphosphoren," *Zeitschrift für Physik*, vol. 94, pp. 38–46, jan 1935.

[6] G. Babitzki, R. Denschlag, and P. Tavan, "Polarization effects stabilize bacteriorhodopsin's chromophore binding pocket: A molecular dynamics study," *The Journal of Physical Chemistry B*, vol. 113, no. 30, pp. 10483–10495, 2009. PMID: 19719289.

[7] J. Leszczynski, *Handbook of computational chemistry*. Springer Science+Business Media B.V, 2012.

[8] R. W. Schoenlein, J. G. Fujimoto, G. L. Eesley, and T. W. Capehart, "Femtosecond relaxation dynamics of image-potential states," *Physical Review B*, vol. 43, pp. 4688–4698, feb 1991.

[9] J. L. Spudich, C.-S. Yang, K.-H. Jung, and E. N. Spudich, "Retinylidene Proteins: Structures and Functions from Archaea to Humans," *Annual Review of Cell and Developmental Biology*, vol. 16, pp. 365–392, nov 2000.

[10] H. G. Khorana, "Two light-transducing membrane proteins: bacteriorhodopsin and the mammalian rhodopsin.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 90, pp. 1166–71, feb 1993.

[11] E. J. Helmreich and K.-P. Hofmann, "Structure and function of proteins in G-protein-coupled signal transfer," *Biochimica et Biophysica Acta (BBA) - Reviews on Biomembranes*, vol. 1286, pp. 285–322, oct 1996.

[12] K. Palczewski, T. Kumasaka, T. Hori, C. A. Behnke, H. Motoshima, B. A. Fox, I. Le Trong, D. C. Teller, T. Okada, R. E. Stenkamp, M. Yamamoto, and M. Miyano, "Crystal structure of rhodopsin: A G protein-coupled receptor.," *Science (New York, N.Y.)*, vol. 289, pp. 739–45, aug 2000.

[13] J. K. McBee, K. Palczewski, W. Baehr, and D. R. Pepperberg, "Confronting Complexity: the Interlink of Phototransduction and Retinoid Metabolism in the Vertebrate Retina," *Progress in Retinal and Eye Research*, vol. 20, pp. 469–529, jul 2001.

[14] K. W. Foster and R. D. Smyth, “Light Antennas in phototactic algae.,” *Microbiological reviews*, vol. 44, pp. 572–630, dec 1980.

[15] K. W. Foster, J. Saranak, N. Patel, G. Zarilli, M. Okabe, T. Kline, and K. Nakanishi, “A rhodopsin is the functional photoreceptor for phototaxis in the unicellular eukaryote Chlamydomonas.,” *Nature*, vol. 311, no. 5988, pp. 756–9.

[16] 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 (New York, N.Y.)*, vol. 296, pp. 2395–8, jun 2002.

[17] 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,” *Proceedings of the National Academy of Sciences*, vol. 100, no. 24, pp. 13940–13945, 2003.

[18] W. Deininger, P. Kröger, U. Hegemann, F. Lottspeich, and P. Hegemann, “Chlamyrhodopsin represents a new type of sensory photoreceptor,” vol. 14, no. 23, pp. 5849–5858, 1995.

[19] S.-I. Ozawa, J. Nield, A. Terao, E. J. Stauber, M. Hippler, H. Koike, J.-D. Rochaix, and Y. Takahashi, “Biochemical and structural studies of the large Ycf4-photosystem I assembly complex of the green alga Chlamydomonas reinhardtii.,” *The Plant cell*, vol. 21, pp. 2424–42, aug 2009.

[20] F. Zhang, M. Prigge, F. Beyrière, S. P. Tsunoda, J. Mattis, O. Yizhar, P. Hegemann, and K. Deisseroth, “Red-shifted optogenetic excitation: a tool for fast neural control derived from Volvox carteri,” *Nature Neuroscience*, vol. 11, pp. 631–633, jun 2008.

[21] E. G. Govorunova, E. N. Spudich, C. E. Lane, O. A. Sineshchekov, and J. L. Spudich, “New channelrhodopsin with a red-shifted spectrum and rapid kinetics from Mesostigma viride.,” *mBio*, vol. 2, pp. e00115–11, jul 2011.

[22] A. Kianianmomeni, K. Stehfest, G. Nematollahi, P. Hegemann, and A. Hallmann, “Channelrhodopsins of Volvox carteri Are Photochromic Proteins That Are Specifically Expressed in Somatic Cells under Control of Light, Temperature, and the Sex Inducer,” *PLANT PHYSIOLOGY*, vol. 151, pp. 347–366, sep 2009.

[23] E. G. Govorunova, O. A. Sineshchekov, H. Li, R. Janz, and J. L. Spudich, “Characterization of a Highly Efficient Blue-shifted Channelrhodopsin from the Marine Alga Platymonas subcordiformis,” *Journal of Biological Chemistry*, vol. 288, pp. 29911–29922, oct 2013.

[24] J. I. Ogren, S. Mamaev, D. Russano, H. Li, J. L. Spudich, and K. J. Rothschild, “Retinal Chromophore Structure and Schiff Base Interactions in Red-Shifted Channelrhodopsin-1 from Chlamydomonas augustae,” *Biochemistry*, vol. 53, pp. 3961–3970, jun 2014.

[25] O. A. Sineshchekov, K.-H. Jung, and J. L. Spudich, “Two rhodopsins mediate phototaxis to low- and high-intensity light in Chlamydomonas reinhardtii.,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, pp. 8689–94, jun 2002.

[26] 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,” *Journal of Biological Chemistry*, vol. 283, no. 50, pp. 35033–35041, 2008.

[27] E. S. Boyden, F. Zhang, E. Bamberg, G. Nagel, and K. Deisseroth, “Millisecond-timescale, genetically targeted optical control of neural activity,” *Nature Neuroscience*, vol. 8, pp. 1263–1268, sep 2005.

[28] 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,” *Proceedings of the National Academy of Sciences*, vol. 102, pp. 17816–17821, dec 2005.

[29] 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,” *Current Biology*, vol. 15, pp. 2279–2284, dec 2005.

[30] 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. Deisseroth, and O. Nureki, “Crystal structure of the channelrhodopsin light-gated cation channel,” *Nature*, vol. 482, no. 7385, pp. 369–374, 2012.

[31] 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 (New York, N.Y.)*, vol. 358, p. eaan8862, nov 2017.

[32] C. Bamann, T. Kirsch, G. Nagel, and E. Bamberg, “Spectral characteristics of the photocycle of channelrhodopsin-2 and its implication for channel function,” *Journal of Molecular Biology*, vol. 375, no. 3, pp. 686 – 694, 2008.

[33] W. Deininger, M. Fuhrmann, and P. Hegemann, “Opsin evolution: out of wild green yonder?,” *Trends in Genetics*, vol. 16, no. 4, pp. 158 – 159, 2000.

[34] 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,” *Journal of Biological Chemistry*, vol. 287, no. 9, pp. 6904–6911, 2012.

[35] O. P. Ernst, P. A. S. Murcia, P. Daldrop, S. P. Tsunoda, S. Kateriya, and P. Hegemann, “Photoactivation of channelrhodopsin,” *Journal of Biological Chemistry*, vol. 283, no. 3, pp. 1637–1643, 2008.

[36] P. Hegemann and G. Nagel, “From channelrhodopsins to optogenetics,” *EMBO Molecular Medicine*, vol. 5, no. 2, pp. 173–176, 2013.

[37] H. C. Watanabe, K. Welke, F. Schneider, S. Tsunoda, F. Zhang, K. Deisseroth, P. Hegemann, and M. Elstner, “Structural model of channelrhodopsin,” *Journal of Biological Chemistry*, vol. 287, no. 10, pp. 7456–7466, 2012.

[38] A. P. Plazzo, N. De Franceschi, F. Da Broi, F. Zonta, M. F. Sanasi, F. Filippini, and M. Mongillo, “Bioinformatic and mutational analysis of channelrhodopsin-2 protein cation-conducting pathway,” *Journal of Biological Chemistry*, vol. 287, no. 7, pp. 4818–4825, 2012.

[39] V. A. Lorenz-Fonfria, 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 channelrhodopsin2 and their relevance to channel gating,” *Proceedings of the National Academy of Sciences*, vol. 110, no. 14, pp. E1273–E1281, 2013.

[40] K. Ruffert, B. Himmel, D. Lall, C. Bamann, E. Bamberg, H. Betz, and V. Eulenburg, “Glutamate residue 90 in the predicted transmembrane domain 2 is crucial for cation flux through channelrhodopsin 2,” *Biochemical and Biophysical Research Communications*, vol. 410, no. 4, pp. 737 – 743, 2011.

[41] Y. Sugiyama, H. Wang, T. Hikima, M. Sato, J. Kuroda, T. Takahashi, T. Ishizuka, and H. Yawo, “Photocurrent attenuation by a single polar-to-nonpolar point mutation of channelrhodopsin-2,” *Photochem. Photobiol. Sci.*, vol. 8, pp. 328–336, 2009.

[42] 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,” *Journal of Molecular Biology*, vol. 425, no. 10, pp. 1795 – 1814, 2013.

[43] M. Nack, I. Radu, C. Bamann, E. Bamberg, and J. Heberle, “The retinal structure of channelrhodopsin-2 assessed by resonance raman spectroscopy,” *{FEBS} Letters*, vol. 583, no. 22, pp. 3676 – 3680, 2009.

[44] S. P. Tsunoda and P. Hegemann, “Glu 87 of channelrhodopsin-1 causes ph-dependent color tuning and fast photocurrent inactivation,” *Photochemistry and Photobiology*, vol. 85, no. 2, pp. 564–569, 2009.

[45] L. Gunaydin, O. Yizhar, A. Berndt, V. Sohal, K. Deisseroth, and P. Hegemann, “Ultrafast optogenetic control,” *Nat. Neuroscience*, vol. 11, pp. 631–633, 2010.

[46] M. K. Verhoefen, C. Bamann, R. Blöcher, U. Förster, E. Bamberg, and J. Wachtveitl, “The photocycle of channelrhodopsin-2: Ultrafast reaction dynamics and subsequent reaction steps,” *ChemPhysChem*, vol. 11, no. 14, pp. 3113–3122, 2010.

[47] J. Kuhne, K. Eisenhauer, E. Ritter, P. Hegemann, K. Gerwert, and F. Bartl, “Early Formation of the Ion-Conducting Pore in Channelrhodopsin-2,” *Angewandte Chemie International Edition*, vol. 54, pp. 4953–4957, apr 2015.

[48] S. Bruun, D. Stoeppler, 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*, vol. 54, pp. 5389–400, sep 2015.

[49] V. A. Lorenz-Fonfria, 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,” *Journal of the American Chemical Society*, vol. 137, no. 5, pp. 1850–1861, 2015. PMID: 25584873.

[50] 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,” 2016.

[51] M. Kamiya, H. E. Kato, R. Ishitani, O. Nureki, and S. Hayashi, “Structural and spectral characterizations of c1c2 channelrhodopsin and its mutants by molecular simulations,” *Chemical Physics Letters*, vol. 556, no. 0, pp. 266 – 271, 2013.

[52] M. Grote, M. Engelhard, and P. Hegemann, “Of ion pumps, sensors and channels perspectives on microbial rhodopsins between science and history,” *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, vol. 1837, no. 5, pp. 533 – 545, 2014. Retinal Proteins.

[53] J. Y. Lin, M. Z. Lin, P. Steinbach, and R. Y. Tsien, “Characterization of engineered channelrhodopsin variants with improved properties and kinetics,” *Biophysical Journal*, vol. 96, no. 5, pp. 1803 – 1814, 2009.

[54] D. Gradmann, A. Berndt, F. Schneider, and P. Hegemann, “Rectification of the channelrhodopsin early conductance,” *Biophysical Journal*, vol. 101, no. 5, pp. 1057 – 1068, 2011.

[55] T. Kottke, V. A. Lórenz-Fonfría, and J. Heberle, “The Grateful Infrared: Sequential Protein Structural Changes Resolved by Infrared Difference Spectroscopy,” *The Journal of Physical Chemistry B*, vol. 121, pp. 335–350, jan 2017.

[56] A. Berndt, O. Yizhar, L. A. Gunaydin, P. Hegemann, and K. Deisseroth, “Bi-stable neural state switches,” *Nature Neuroscience*, vol. 12, pp. 229–234, feb 2009.

[57] O. Yizhar, L. E. Fenno, M. Prigge, F. Schneider, T. J. Davidson, D. J. O’Shea, V. S. Sohal, I. Goshen, J. Finkelstein, J. T. Paz, K. Stehfest, R. Fudim, C. Ramakrishnan, J. R. Huguenard, P. Hegemann, and K. Deisseroth, “Neocortical excitation/inhibition balance in information processing and social dysfunction,” *Nature*, vol. 477, pp. 171–178, sep 2011.

[58] I. Radu, C. Bamann, M. Nack, G. Nagel, E. Bamberg, and J. Heberle, “Conformational Changes of Channelrhodopsin-2,” *Journal of the American Chemical Society*, vol. 131, pp. 7313–7319, jun 2009.

[59] K. Stehfest, E. Ritter, A. Berndt, F. Bartl, and P. Hegemann, “The branched photocycle of the slow-cycling channelrhodopsin-2 mutant {C128T},” *Journal of Molecular Biology*, vol. 398, no. 5, pp. 690 – 702, 2010.

[60] V. A. Lorenz-Fonfria and J. Heberle, “Channelrhodopsin unchained: Structure and mechanism of a light-gated cation channel,” *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, vol. 1837, no. 5, pp. 626 – 642, 2014. Retinal Proteins.

[61] 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.,” *Journal of the American Chemical Society*, vol. 137, no. 5, pp. 1850–61, 2015.

[62] F. Schneider, C. Grimm, and P. Hegemann, “Biophysics of Channelrhodopsin.,” *Annual review of biophysics*, vol. 44, pp. 167–186, 2015.

[63] C. Bamann, R. Gueta, S. Kleinlogel, G. Nagel, and E. Bamberg, “Structural guidance of the photocycle of channelrhodopsin-2 by an interhelical hydrogen bond,” *Biochemistry*, vol. 49, no. 2, pp. 267–278, 2010. PMID: 20000562.

[64] M. Nack, I. Radu, B.-J. Schultz, T. Resler, R. Schlesinger, A.-N. Bondar, C. del Val, S. Abbruzzetti, C. Viappiani, C. Bamann, E. Bamberg, and J. Heberle, “Kinetics of proton release and uptake by channelrhodopsin-2,” *FEBS Letters*, vol. 586, no. 9, pp. 1344 – 1348, 2012.

[65] A. Berndt, P. Schoenenberger, J. Mattis, K. M. Tye, K. Deisseroth, P. Hegemann, and T. G. Oertner, “High-efficiency channelrhodopsins for fast neuronal stimulation at low light levels,” *Proceedings of the National Academy of Sciences*, vol. 108, pp. 7595–7600, may 2011.

[66] S. Kateriya, G. Nagel, E. Bamberg, and P. Hegemann, “"Vision" in Single-Celled Algae,” *Physiology*, vol. 19, pp. 133–137, jun 2004.

[67] M. Luck and P. Hegemann, “The two parallel photocycles of the Chlamydomonas sensory photoreceptor histidine kinase rhodopsin 1,” *Journal of Plant Physiology*, vol. 217, pp. 77–84, oct 2017.

[68] M. Luck, T. Mathes, S. Bruun, R. Fudim, R. Hagedorn, T. M. Tran Nguyen, S. Kateriya, J. T. M. Kennis, P. Hildebrandt, and P. Hegemann, “A photochromic histidine kinase rhodopsin (HKR1) that is bimodally switched by ultraviolet and blue light.,” *The Journal of biological chemistry*, vol. 287, no. 47, pp. 40083–40090, 2012.

[69] M. Luck, S. Bruun, A. Keidel, P. Hegemann, and P. Hildebrandt, “Photochemical chromophore isomerization in histidine kinase rhodopsin HKR1.,” *FEBS letters*, vol. 589, pp. 1067–71, apr 2015.

[70] F. Jensen, *Introduction to Computational Chemistry*. Chichester: John Wiley and Sons Ltd, 2 ed., 2007.

[71] C. J. Cramer, *Essentials of Computational Chemistry Theories and Models*. Wiley, second edition ed., 2004.

[72] W. Koch and M. C. Holthausen, *A chemist's guide to density functional theory*. Wiley-VCH, 2001.

[73] R. G. Parr and W. Yang, *Density-functional theory of atoms and molecules*. Oxford University Press, 1989.

[74] P.-Å. Malmqvist and B. O. Roos, “The CASSCF state interaction method,” *Chemical Physics Letters*, vol. 155, pp. 189–194, feb 1989.

[75] P. Hohenberg and W. Kohn, “Inhomogeneous electron gas,” *Phys. Rev.*, vol. 136, pp. B864–B871, Nov 1964.

[76] W. Kohn and L. J. Sham, “Self-consistent equations including exchange and correlation effects,” *Phys. Rev.*, vol. 140, pp. A1133–A1138, Nov 1965.

[77] J. P. Perdew, K. Burke, and M. Ernzerhof, “Generalized gradient approximation made simple,” *Phys. Rev. Lett.*, vol. 77, pp. 3865–3868, Oct 1996.

[78] 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. B*, vol. 37, pp. 785–789, Jan 1988.

[79] A. D. Becke, “Density-functional exchange-energy approximation with correct asymptotic behavior,” *Phys. Rev. A*, vol. 38, pp. 3098–3100, Sep 1988.

[80] A. D. Becke, “Density - functional thermochemistry. iii. the role of exact exchange,” *The Journal of Chemical Physics*, vol. 98, no. 7, pp. 5648–5652, 1993.

[81] E. Runge and E. K. U. Gross, “Density-Functional Theory for Time-Dependent Systems,” *Physical Review Letters*, vol. 52, pp. 997–1000, mar 1984.

[82] J. C. Tully, “Molecular dynamics with electronic transitions,” *The Journal of Chemical Physics*, vol. 93, pp. 1061–1071, jul 1990.

[83] S. Hammes?Schiffer and J. C. Tully, “Proton transfer in solution: Molecular dynamics with quantum transitions,” *The Journal of Chemical Physics*, vol. 101, pp. 4657–4667, sep 1994.

[84] C. Zener, “Non-Adiabatic Crossing of Energy Levels,” *Proceedings of the Royal Society of London Series A*, vol. 137, pp. 696–702, Sept. 1932.

[85] L. Landau, “Zur Theorie der Energieubertragung II,” *Physikalische Zeitung der Sowjetunion*, vol. 46, pp. 1–13, Sept. 1932.

[86] M. Gaus, *Extension and Parametrization of an Approximate Density Functional Method for Organic Biomolecules*. PhD thesis, Karlsruhe Institute of Technology, 2011.

[87] 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*, vol. 58, pp. 7260–7268, Sep 1998.

[88] M. Gaus, Q. Cui, and M. Elstner, “DFTB3: Extension of the Self-Consistent-Charge Density-Functional Tight-Binding Method (SCC-DFTB),” *Journal of Chemical Theory and Computation*, vol. 7, pp. 931–948, apr 2011.

[89] D. Porezag, T. Frauenheim, T. Köhler, G. Seifert, and R. Kaschner, “Construction of tight-binding-like potentials on the basis of density-functional theory: Application to carbon,” *Phys. Rev. B*, vol. 51, pp. 12947–12957, May 1995.

[90] G. Seifert and J.-O. Joswig, “Density-functional tight binding - an approximate density-functional theory method,” *Wiley Interdisciplinary Reviews: Computational Molecular Science*, vol. 2, no. 3, pp. 456–465, 2012.

[91] W. Weber and W. Thiel, “Orthogonalization corrections for semiempirical methods,” *Theoretical Chemistry Accounts: Theory, Computation, and Modeling (Theoretica Chimica Acta)*, vol. 103, pp. 495–506, apr 2000.

[92] J. V. Neumann, “Proof of the Quasi-Ergodic Hypothesis.,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 18, pp. 70–82, jan 1932.

[93] J. V. Neumann, “Physical Applications of the Ergodic Hypothesis.,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 18, pp. 263–6, mar 1932.

[94] G. D. Birkhoff, “Proof of the Ergodic Theorem.,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 17, pp. 656–60, dec 1931.

[95] A. Laio and M. Parrinello, “Escaping free-energy minima.,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, pp. 12562–6, oct 2002.

[96] M. Duñach, T. Marti, H. G. Khorana, and K. J. Rothschild, “Uv-visible spectroscopy of bacteriorhodopsin mutants: substitution of Arg-82, Asp-85, Tyr-185, and Asp-212 results in abnormal light-dark adaptation.,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 87, pp. 9873–7, dec 1990.

[97] G. F. Schertler, R. Lozier, H. Michel, and D. Oesterhelt, “Chromophore motion during the bacteriorhodopsin photocycle: polarized absorption spectroscopy of bacteriorhodopsin and its M-state in bacteriorhodopsin crystals.,” *The EMBO journal*, vol. 10, pp. 2353–61, sep 1991.

[98] K. Welke, *QM/MM Simulations of Channelrhodopsins - Elucidating Structure and Spectroscopic Properties*. PhD thesis, Karlsuhe Institute of Technology, 2013. Karlsruhe, KIT, Diss., 2013.

[99] K. Welke, J. S. Frähmcke, H. C. Watanabe, P. Hegemann, and M. Elstner, “Color tuning in binding pocket models of the chlamydomonas-type channelrhodopsins,” *The Journal of Physical Chemistry B*, vol. 115, no. 50, pp. 15119–15128, 2011. PMID: 22077286.

[100] M.-K. Neumann-Verhoefen, 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," *Journal of the American Chemical Society*, vol. 135, no. 18, pp. 6968–6976, 2013.

[101] 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.," *Photochemical & photobiological sciences : Official journal of the European Photochemistry Association and the European Society for Photobiology*, vol. 9, pp. 194–8, feb 2010.

[102] 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.*, vol. 15, pp. 6651–6659, 2013.

[103] A. Ardevol and G. Hummer, "Retinal isomerization and water-pore formation in channelrhodopsin-2.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 115, pp. 3557–3562, apr 2018.

[104] L. Zhang and J. Hermans, "Hydrophilicity of cavities in proteins," *Proteins: Structure, Function, and Genetics*, vol. 24, pp. 433–438, apr 1996.

[105] M. Mueller, C. Bamann, E. Bamberg, and W. Kuehlbrandt, "Projection structure of channelrhodopsin-2 at 6 åresolution by electron crystallography," *Journal of Molecular Biology*, vol. 414, no. 1, pp. 86 – 95, 2011.

[106] F. Beyle, *Description of the Channelrhodopsin Active Site: a Computational Model*. PhD thesis, Karlsuhe Institute of Technology, 2014. Karlsruhe, KIT, Master thesis, 2014.

[107] 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," *The Journal of Physical Chemistry B*, vol. 114, no. 23, pp. 7830–7843, 2010.

[108] T. J. Piggot, n. Piñeiro, and S. Khalid, "Molecular dynamics simulations of phosphatidylcholine membranes: A comparative force field study," *Journal of Chemical Theory and Computation*, vol. 8, no. 11, pp. 4593–4609, 2012.

[109] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, "Comparison of simple potential functions for simulating liquid water," *The Journal of Chemical Physics*, vol. 79, no. 2, pp. 926–935, 1983.

[110] D. der Spoel, E.Lindahl, B.Hess, A. van Buuren, E. Apol, P. Meulenhoff, D. Tielemann, A. Sijbers, K. Feenstra, R. van Drunen, and H. Berendsen, "www.gromacs.org," *Gromacs User Manual Version 4.5.4*, 2010.

[111] E. Lindahl, B. Hess, and D. van der Spoel, "Gromacs 3.0: a package for molecular simulation and trajectory analysis," *Molecular modeling annual*, vol. 7, no. 8, pp. 306–317, 2001.

[112] K. Gerwert, E. Freier, and S. Wolf, "The role of protein-bound water molecules in microbial rhodopsins," *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, vol. 1837, pp. 606–613, may 2014.

[113] T. Wolter, M. Elstner, S. Fischer, J. C. Smith, and A.-N. Bondar, "Mechanism by which Untwisting of Retinal Leads to Productive," *The journal of physical chemistry. B*, vol. 119, pp. 2229–2240, 2015.

[114] 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," *The Journal of Chemical Physics*, vol. 110, p. 10452, may 1999.

[115] 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.*, mar 2016.

[116] M. Gaus, Q. Cui, and M. Elstner, "Density functional tight binding: application to organic and biological molecules," *Wiley Interdisciplinary Reviews: Computational Molecular Science*, vol. 4, no. 1, pp. 49–61, 2014.

[117] M. Gaus, A. Goez, and M. Elstner, "Parametrization and Benchmark of DFTB3 for Organic Molecules," *Journal of Chemical Theory and Computation*, vol. 9, pp. 338–354, jan 2013.

[118] M. Gaus, A. Goez, and M. Elstner, "Parametrization and benchmark of dftb3 for organic molecules," *Journal of Chemical Theory and Computation*, vol. 9, no. 1, pp. 338–354, 2013.

[119] 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.*, vol. 15, pp. 12582–12590, 2013.

[120] W. B. Fischer, S. Sonar, T. Marti, H. G. Khorana, and K. J. Rothschild, "Detection of a water molecule in the active-site of bacteriorhodopsin: hydrogen bonding changes during the primary photoreaction," *Biochemistry*, vol. 33, pp. 12757–12762, jan 1994.

[121] Y. V. Guo, *Theoretical investigation of Kainate Receptor GluK2 and Channelrhodopsin-2 : structure and mechanism*. PhD thesis, Karlsruhe, [2017].

[122] W. Thiel, "MNDO99 Program Package," vol. Version 6., no. Germany, pp. Max–PLanck–Institute für Kohlenforschung Mülheim a, 2006.

[123] A. Barducci, G. Bussi, and M. Parrinello, "Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method," *Physical Review Letters*, vol. 100, p. 020603, jan 2008.

[124] A. Laio and F. L. Gervasio, "Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science," *Reports on Progress in Physics*, vol. 71, p. 126601, dec 2008.

[125] A. Barducci, M. Bonomi, and M. Parrinello, "Metadynamics," *Wiley Interdisciplinary Reviews: Computational Molecular Science*, vol. 1, pp. 826–843, sep 2011.

[126] J. C. Lansing, M. Hohwy, C. P. Jaroniec, A. F. L. Creemers, J. Lugtenburg, J. Herzfeld, and R. G. Griffin, "Chromophore distortions in the bacteriorhodopsin photocycle: evolution of the H-C14-C15-H dihedral angle measured by solid-state NMR.," *Biochemistry*, vol. 41, pp. 431–8, jan 2002.

[127] J. Mao, N.-N. Do, F. Scholz, L. Reggie, M. Mehler, A. Lakatos, Y.-S. Ong, S. J. Ullrich, L. J. Brown, R. C. D. Brown, J. Becker-Baldus, J. Wachtveitl, and C. Glaubitz, "Structural Basis of the

Green-Blue Color Switching in Proteorhodopsin as Determined by NMR Spectroscopy," *Journal of the American Chemical Society*, vol. 136, pp. 17578–17590, dec 2014.

[128] H. Luecke, B. Schobert, H. T. Richter, J. P. Cartailler, and J. K. Lanyi, "Structure of bacteriorhodopsin at 1.55 A resolution.," *Journal of molecular biology*, vol. 291, pp. 899–911, aug 1999.

[129] T. Nishikawa, M. Murakami, and T. Kouyama, "Crystal structure of the 13-cis isomer of bacteriorhodopsin in the dark-adapted state.," *Journal of molecular biology*, vol. 352, pp. 319–28, sep 2005.

[130] T. Ran, G. Ozorowski, Y. Gao, O. A. Sineshchekov, W. Wang, J. L. Spudich, H. Luecke, and IUCr, "Cross-protomer interaction with the photoactive site in oligomeric proteorhodopsin complexes," *Acta Crystallographica Section D Biological Crystallography*, vol. 69, pp. 1965–1980, oct 2013.

[131] 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," *Journal of the American Chemical Society*, vol. 136, no. 9, pp. 3475–3482, 2014.

[132] M. R. VanGordon, G. Gyawali, S. W. Rick, and S. B. Rempe, "Atomistic Study of Intramolecular Interactions in the Closed-State Channelrhodopsin Chimera, C1C2," *Biophysical Journal*, vol. 112, no. 5, pp. 943–952, 2017.

[133] M. Biasini, S. Bienert, A. Waterhouse, K. Arnold, G. Studer, T. Schmidt, F. Kiefer, T. G. Cassarino, M. Bertoni, L. Bordoli, and T. Schwede, "SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information," *Nucleic Acids Research*, vol. 42, pp. W252–W258, jul 2014.

[134] K. Arnold, L. Bordoli, J. Kopp, and T. Schwede, "The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling," *Bioinformatics*, vol. 22, pp. 195–201, jan 2006.

[135] N. Guex, M. C. Peitsch, and T. Schwede, "Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective," *ELECTROPHORESIS*, vol. 30, pp. S162–S173, jun 2009.

[136] L. Vogeley, O. A. Sineshchekov, V. D. Trivedi, J. Sasaki, J. L. Spudich, and H. Luecke, "Anabaena sensory rhodopsin: a photochromic color sensor at 2.0 A.," *Science (New York, N.Y.)*, vol. 306, pp. 1390–3, nov 2004.

[137] T. Nakanishi, S. Kanada, M. Murakami, K. Ihara, and T. Kouyama, "Large deformation of helix F during the photoreaction cycle of Pharaonis halorhodopsin in complex with azide.," *Biophysical journal*, vol. 104, pp. 377–85, jan 2013.

[138] A. Gautier, H. R. Mott, M. J. Bostock, J. P. Kirkpatrick, and D. Nietlispach, "Structure determination of the seven-helix transmembrane receptor sensory rhodopsin II by solution NMR spectroscopy.," *Nature structural & molecular biology*, vol. 17, pp. 768–74, jun 2010.

[139] S. Faham, D. Yang, E. Bare, S. Yohannan, J. P. Whitelegge, and J. U. Bowie, "Side-chain contributions to membrane protein structure and stability.," *Journal of molecular biology*, vol. 335, pp. 297–305, jan 2004.

[140] M. Furuse, J. Tamogami, T. Hosaka, T. Kikukawa, N. Shinya, M. Hato, N. Ohsawa, S. Y. Kim, K. H. Jung, M. Demura, S. Miyauchi, N. Kamo, K. Shimono, T. Kimura-Someya, S. Yokoyama, and M. Shirouzu, "Structural basis for the slow photocycle and late proton release in Acetabularia rhodopsin I from the marine plant Acetabularia acetabulum.," *Acta crystallographica. Section D, Biological crystallography*, vol. 71, pp. 2203–16, nov 2015.

[141] L. A. Kelley, S. Mezulis, C. M. Yates, M. N. Wass, and M. J. E. Sternberg, "The Phyre2 web portal for protein modeling, prediction and analysis," *Nature Protocols*, vol. 10, pp. 845–858, may 2015.

[142] M. Feig, J. Karanicolas, and C. L. Brooks, "MMTSB Tool Set: enhanced sampling and multiscale modeling methods for applications in structural biology," *Journal of Molecular Graphics and Modelling*, vol. 22, pp. 377–395, may 2004.

[143] 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," *Journal of the American Chemical Society*, vol. 128, no. 33, pp. 10808–10818, 2006. PMID: 16910676.

[144] 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," *Journal of Physical Chemistry B*, vol. 109, no. 8, pp. 3606–3615, 2005. cited By (since 1996)148.

[145] V. Veryazov, P.-O. Widmark, L. Serrano-Andrés, R. Lindh, and B. O. Roos, "2MOLCAS as a development platform for quantum chemistry software," *International Journal of Quantum Chemistry*, vol. 100, no. 4, pp. 626–635, 2004.

[146] F. Aquilante, J. Autschbach, R. K. Carlson, L. F. Chibotaru, M. G. Delcey, L. De Vico, I. Fdez. Galván, N. Ferré, L. M. Frutos, L. Gagliardi, M. Garavelli, A. Giussani, C. E. Hoyer, G. Li Manni, H. Lischka, D. Ma, P. Å. Malmqvist, T. Müller, A. Nenov, M. Olivucci, T. B. Pedersen, D. Peng, F. Plasser, B. Pritchard, M. Reiher, I. Rivalta, I. Schapiro, J. Segarra-Martí, M. Stenrup, D. G. Truhlar, L. Ungur, A. Valentini, S. Vancoillie, V. Veryazov, V. P. Vysotskiy, O. Weingart, F. Zapata, and R. Lindh, "Molcas 8: New capabilities for multiconfigurational quantum chemical calculations across the periodic table," *Journal of Computational Chemistry*, vol. 37, pp. 506–541, feb 2016.

[147] J. W. Ponder and F. M. Richards, "An efficient newton-like method for molecular mechanics energy minimization of large molecules," *Journal of Computational Chemistry*, vol. 8, pp. 1016–1024, oct 1987.

[148] A. Penzkofer, M. Luck, T. Mathes, and P. Hegemann, "Bistable retinal schiff base photodynamics of histidine kinase rhodopsin HKR1 from Chlamydomonas reinhardtii.," *Photochemistry and photobiology*, vol. 90, pp. 773–85, jan 2014.

[149] C. S. Page and M. Olivucci, "Ground and excited state CASPT2 geometry optimizations of small organic molecules," *Journal of Computational Chemistry*, vol. 24, pp. 298–309, feb 2003.

[150] R. González-Luque, M. Garavelli, F. Bernardi, M. Merchán, M. A. Robb, and M. Olivucci, "Computational evidence in favor of a two-state, two-mode model of the retinal chromophore photoisomerization.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 97, pp. 9379–84, aug 2000.

[151] T. W. Keal, M. Wanko, and W. Thiel, “Assessment of semiempirical methods for the photoisomerisation of a protonated Schiff base,” *Theoretical Chemistry Accounts*, vol. 123, pp. 145–156, may 2009.

[152] S. Gozem, M. Huntress, I. Schapiro, R. Lindh, A. A. Granovsky, C. Angeli, and M. Olivucci, “Dynamic electron correlation effects on the ground state potential energy surface of a retinal chromophore model,” *Journal of Chemical Theory and Computation*, vol. 8, no. 11, pp. 4069–4080, 2012.

[153] S. Gozem, F. Melaccio, R. Lindh, A. I. Krylov, A. A. Granovsky, C. Angeli, and M. Olivucci, “Mapping the Excited State Potential Energy Surface of a Retinal Chromophore Model with Multireference and Equation-of-Motion Coupled-Cluster Methods,” *Journal of Chemical Theory and Computation*, vol. 9, pp. 4495–4506, oct 2013.

[154] I. Dokukina, C. M. Marian, and O. Weingart, “New Perspectives on an Old Issue: A Comparative MS-CASPT2 and OM2-MRCI Study of Polyenes and Protonated Schiff Bases,” *Photochemistry and Photobiology*, vol. 93, pp. 1345–1355, nov 2017.

[155] T. Kobayashi, T. Saito, and H. Ohtani, “Real-time spectroscopy of transition states in bacteriorhodopsin during retinal isomerization,” *Nature*, vol. 414, pp. 531–534, nov 2001.

[156] P. Hegemann, M. Fuhrmann, and S. Kateriya, “ALGAL SENSORY PHOTORECEPTORS,” *Journal of Phycology*, vol. 37, pp. 668–676, oct 2001.

[157] . M. Garavelli, . P. Celani, . F. Bernardi, *, , . M. A. Robb, *, and . M. Olivucci*, “The C5H6NH2+ Protonated Shiff Base:? An ab Initio Minimal Model for Retinal Photoisomerization,” 1997.

[158] M. Garavelli, F. Bernardi, M. Olivucci, T. Vreven, S. Klein, P. Celani, and M. A. Robb, “Potential-energy surfaces for ultrafast photochemistry Static and dynamic aspects,” *Faraday Discussions*, vol. 110, pp. 51–70, jan 1998.

[159] Oliver Weingart, Annapaola Migani, Massimo Olivucci , Michael A. Robb , and Volker Buss, and P. Hunt, “Probing the Photochemical Funnel of a Retinal Chromophore Model via Zero-Point Energy Sampling Semiclassical Dynamics,” 2004.

[160] O. Weingart, V. Buss, and M. Robb, “Excited state molecular dynamics of retinal model chromophores,” *Phase Transitions*, vol. 78, pp. 17–24, jan 2005.

[161] B. G. Levine and T. J. Martínez, “Isomerization Through Conical Intersections,” *Annual Review of Physical Chemistry*, vol. 58, pp. 613–634, may 2007.

[162] J. J. Szymczak, M. Barbatti, and H. Lischka, “Mechanism of Ultrafast Photodecay in Restricted Motions in Protonated Schiff Bases: The Pentadieniminium Cation,” *Journal of Chemical Theory and Computation*, vol. 4, pp. 1189–1199, aug 2008.

[163] M. Barbatti, M. Ruckenbauer, J. J. Szymczak, A. J. A. Aquino, and H. Lischka, “Nonadiabatic excited-state dynamics of polar π-systems and related model compounds of biological relevance,” *Phys. Chem. Chem. Phys.*, vol. 10, no. 4, pp. 482–494, 2008.

[164] S. Yang and T. J. Martínez, “Ab InitioMultiple Spawning: First Principles Dynamics Around Conical Intersections,” pp. 347–374, nov 2011.

[165] N. Klaffki, O. Weingart, M. Garavelli, and E. Spohr, “Sampling excited state dynamics: influence of HOOP mode excitations in a retinal model,” *Physical Chemistry Chemical Physics*, vol. 14, p. 14299, oct 2012.

[166] X. Xu, S. Gozem, M. Olivucci, and D. G. Truhlar, “Combined self-consistent-field and spin-flip tamm-dancoff density functional approach to potential energy surfaces for photochemistry,” *Journal of Physical Chemistry Letters*, vol. 4, no. 2, pp. 253–258, 2013.

[167] M. Huix-Rotllant, M. Filatov, S. Gozem, I. Schapiro, M. Olivucci, and N. Ferré, “Assessment of Density Functional Theory for Describing the Correlation Effects on the Ground and Excited State Potential Energy Surfaces of a Retinal Chromophore Model,” *Journal of Chemical Theory and Computation*, vol. 9, pp. 3917–3932, jan 2013.

[168] S. Gozem, A. I. Krylov, and M. Olivucci, “Conical intersection and potential energy surface features of a model retinal chromophore: Comparison of EOM-CC and multireference methods,” *Journal of Chemical Theory and Computation*, vol. 9, no. 1, pp. 284–292, 2013.

[169] I. Schapiro and F. Neese, “SORCI for photochemical and thermal reaction paths: A benchmark study,” *Computational and Theoretical Chemistry*, vol. 1040-1041, no. 0, pp. 84–98, 2014.

[170] S. Gozem, F. Melaccio, A. Valentini, M. Filatov, M. Huix-Rotllant, N. Ferré, L. M. Frutos, C. Angeli, A. I. Krylov, A. A. Granovsky, R. Lindh, and M. Olivucci, “Shape of Multireference, Equation-of-Motion Coupled-Cluster, and Density Functional Theory Potential Energy Surfaces at a Conical Intersection.,” *Journal of chemical theory and computation*, vol. 10, pp. 3074–3084, aug 2014.

[171] A. Zen, E. Coccia, S. Gozem, M. Olivucci, and L. Guidoni, “Quantum monte carlo treatment of the charge transfer and diradical electronic character in a retinal chromophore minimal model,” *Journal of Chemical Theory and Computation*, vol. 11, no. 3, pp. 992–1005, 2015.

[172] D. Tuna, Y. Lu, A. Koslowski, and W. Thiel, “Semiempirical Quantum-Chemical Orthogonalization-Corrected Methods: Benchmarks of Electronically Excited States,” *Journal of Chemical Theory and Computation*, vol. 12, pp. 4400–4422, sep 2016.

[173] M. Manathunga, X. Yang, H. L. Luk, S. Gozem, L. M. Frutos, A. Valentini, N. Ferrè, and M. Olivucci, “Probing the Photodynamics of Rhodopsins with Reduced Retinal Chromophores,” *Journal of Chemical Theory and Computation*, vol. 12, pp. 839–850, feb 2016.

[174] Simona Fantacci, Annapaola Migani, and Massimo Olivucci, “CASPT2//CASSCF and TDDFT//CASSCF Mapping of the Excited State Isomerization Path of a Minimal Model of the Retinal Chromophore,” 2004.

[175] L. Liu, J. Liu, and T. J. Martinez, “Dynamical Correlation Effects on Photoisomerization: Ab Initio Multiple Spawning Dynamics with MS-CASPT2 for a Model trans-Protonated Schiff Base,” *The Journal of Physical Chemistry B*, vol. 120, pp. 1940–1949, mar 2016.

[176] K. Andersson, P. Malmqvist, and B. O. Roos, “Second?order perturbation theory with a complete active space self?consistent field reference function,” *The Journal of Chemical Physics*, vol. 96, pp. 1218–1226, jan 1992.

[177] J. Finley, P.-Å. Malmqvist, B. O. Roos, and L. Serrano-Andrés, “The multi-state CASPT2 method,” *Chemical Physics Letters*, vol. 288, pp. 299–306, may 1998.

[178] A. A. Granovsky, “Extended multi-configuration quasi-degenerate perturbation theory: The new approach to multi-state multi-reference perturbation theory,” *The Journal of Chemical Physics*, vol. 134, p. 214113, jun 2011.

[179] T. Shiozaki, W. Gy?rffy, P. Celani, and H.-J. Werner, "Communication: Extended multi-state complete active space second-order perturbation theory: Energy and nuclear gradients," *The Journal of Chemical Physics*, vol. 135, p. 081106, aug 2011.

[180] B. Vlaisavljevich and T. Shiozaki, "Nuclear Energy Gradients for Internally Contracted Complete Active Space Second-Order Perturbation Theory: Multistate Extensions," *Journal of Chemical Theory and Computation*, vol. 12, pp. 3781–3787, aug 2016.

[181] M. K. MacLeod and T. Shiozaki, "Communication: Automatic code generation enables nuclear gradient computations for fully internally contracted multireference theory," *The Journal of Chemical Physics*, vol. 142, p. 051103, feb 2015.

[182] J. W. Park and T. Shiozaki, "On-the-Fly CASPT2 Surface-Hopping Dynamics," *Journal of Chemical Theory and Computation*, vol. 13, pp. 3676–3683, aug 2017.

[183] A. Koslowski, 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," *Journal of Computational Chemistry*, vol. 24, no. 6, pp. 714–726, 2003.

[184] W. Thiel, "MNDO program," vol. Version 7, no. Germany, pp. Max–Planck–Institut für Kohlenforschung: Mülheim a, 2017.

[185] T. W. Keal, A. Koslowski, and W. Thiel, "Comparison of algorithms for conical intersection optimisation using semiempirical methods," *Theoretical Chemistry Accounts*, vol. 118, pp. 837–844, nov 2007.

[186] L. Spörkel and W. Thiel, "Adaptive time steps in trajectory surface hopping simulations," *The Journal of Chemical Physics*, vol. 144, p. 194108, may 2016.

[187] "Gaussian 09, revision a.02, gaussian, inc., wallingford ct, 2016."

[188] O. N. Burchak, L. Mugherli, M. Ostuni, J. J. Lacap?re, and M. Y. Balakirev, "Combinatorial discovery of fluorescent pharmacophores by multicomponent reactions in droplet arrays," *Journal of the American Chemical Society*, vol. 133, pp. 10058–10061, jul 2011.

[189] A. Köhn and C. Hättig, "Analytic gradients for excited states in the coupled-cluster model CC2 employing the resolution-of-the-identity approximation," *The Journal of Chemical Physics*, vol. 119, pp. 5021–5036, sep 2003.

[190] C. Hättig and A. Köhn, "Transition moments and excited-state first-order properties in the coupled-cluster model CC2 using the resolution-of-the-identity approximation," *The Journal of Chemical Physics*, vol. 117, pp. 6939–6951, oct 2002.

[191] C. Hättig, "Geometry optimizations with the coupled-cluster model CC2 using the resolution-of-the-identity approximation," *The Journal of Chemical Physics*, vol. 118, pp. 7751–7761, may 2003.

[192] "TURBOMOLE V6.2 2010, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989-2007, TURBOMOLE GmbH, since 2007; available from `http://www.turbomole.com`."

[193] K. E. Yousaf and K. A. Peterson, "Optimized auxiliary basis sets for explicitly correlated methods," *The Journal of Chemical Physics*, vol. 129, p. 184108, nov 2008.

[194] L. I. Dixon, M. A. Carroll, T. J. Gregson, G. J. Ellames, R. W. Harrington, and W. Clegg, "Unprecedented regiochemical control in the formation of aryl[1,2-a]imidazopyridines from alkynyliodonium salts: mechanistic insights," *Org. Biomol. Chem.*, vol. 11, pp. 5877–5884, 2013.

[195] F. Neese, "The orca program system," *Wiley Interdisciplinary Reviews: Computational Molecular Science*, vol. 2, no. 1, pp. 73–78, 2012.

[196] e. a. D.A. Case, J.T. Berryman, "Amber 2015, university of california, san francisco," 2015.

[197] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case, "Development and testing of a general amber force field," *Journal of Computational Chemistry*, vol. 25, pp. 1157–1174, jul 2004.

[198] C. I. Bayly, P. Cieplak, W. Cornell, and P. A. Kollman, "A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model," *The Journal of Physical Chemistry*, vol. 97, pp. 10269–10280, oct 1993.

[199] T. Yanai, D. P. Tew, and N. C. Handy, "A new hybrid exchange-correlation functional using the Coulomb-attenuating method (CAM-B3LYP)," *Chemical Physics Letters*, vol. 393, pp. 51–57, jul 2004.

[200] KOALA, an ab-initio electronic structure program, written by S. Höfener, with contributions from A.-S. Hehn, J. Heuser and N. Schieschke.

[201] S. Höfener, "Coupled-cluster frozen-density embedding using resolution of the identity methods," *J. Comput. Chem.*, vol. 35, no. 23, pp. 1716–1724, 2014.

[202] 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.*, vol. 4, pp. 435–447, Feb. 2008.

[203] M. Gaus, X. Lu, M. Elstner, and Q. Cui, "Parameterization of DFTB3/3OB for Sulfur and Phosphorus for Chemical and Biological Applications," *Journal of Chemical Theory and Computation*, vol. 10, pp. 1518–1537, apr 2014.

[204] B. Aradi, B. Hourahine, and T. Frauenheim, "DFTB+, a Sparse Matrix-Based Implementation of the DFTB Method," 2007.

[205] 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," *Journal of Chemical Theory and Computation*, vol. 13, pp. 1737–1747, apr 2017.

[206] F. H. Allen, O. Kennard, D. G. Watson, L. Brammer, A. G. Orpen, and R. Taylor, "Tables of bond lengths determined by X-ray and neutron diffraction. Part 1. Bond lengths in organic compounds," *Journal of the Chemical Society, Perkin Transactions 2*, vol. 0, p. S1, jan 1987.

[207] Marcin Hoffmann, Agnieszka Plutecka, Urszula Rychlewska, Zdzislaw Kucybala, and J. P. Ilona, "New Type of Bonding Formed from an Overlap between π Aromatic and π^* CO Molecular Orbitals Stabilizes the Coexistence in One Molecule of the Ionic and Neutral meso-Ionic Forms of Imidazopyridine," 2005.

[208] F. Neese, "The ORCA program system," *Wiley Interdisciplinary Reviews: Computational Molecular Science*, vol. 2, pp. 73–78, jan 2012.

[209] F. Neese, “Software update: the ORCA program system, version 4.0,” *Wiley Interdisciplinary Reviews: Computational Molecular Science*, vol. 8, p. e1327, jan 2018.

Part IV

Appendix

A | Dark state of the Channelrhodopsin

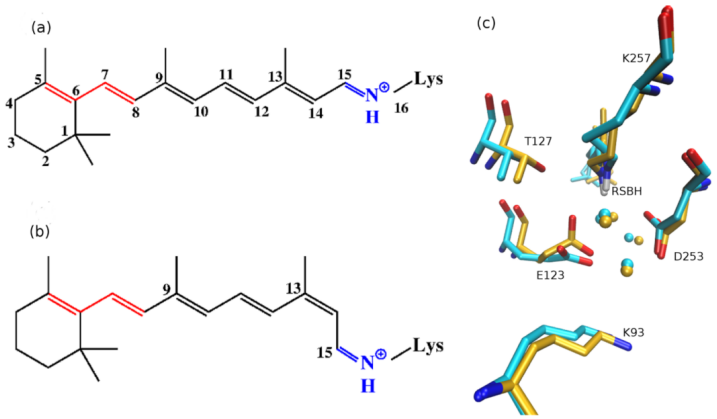


Figure 1: a) chemical structure of all-*trans* retinal and b) of 13-*cis*, 15-*syn* and in c) aligned structure of active site with all-*trans* (cyan backbone) and 13-*cis*, 15-*syn* (yellow backbone) retinal.

Table 1: Summary of structural motifs of the 14 cooling QM/MM trajectories at the beginning of the trajectory at 300 K and at the end of the trajectory at 100 K.

Trajectory no.	300K	100K
1	E123 up[a]	E123 up
2	E123 up	E123 up
3	E123 up	E123 up
4	E123 up	E123 up
5	H2O up[b]	E123 up
6	H2O up	E123 up
7	D253 up[c]	E123 up
8	D253 up	E123 up
9	D253 up	E123 up
10	H2O up	H2O up
		E123 up
11	D253 down[d]	D253 down
12	D253 down	D253 down
13	H2O down[e]	H2O down
		D253 down
14	H2O down	H2O down

[a] E123 up: $RSBH^+$...O-E123; E123-upward
[b] H2O up: $RSBH^+$...OH_2; E123-upward
[c] D253 up:-$RSBH^+$...O-D253; E123-upward
[d] D253 down: $RSBH^+$...O-D253; E123-downward
[e] H2O down: $RSBH^+$...OH_2; E123-downward

Table 2: Point in simulation when E90 flips down. Four 1 µs long MD simulations have been performed

	point when E90 flip down	
MD1	Monomer 1	27 ns
	Monomer 2	600 ns
MD2	Monomer 1	500 ns
	Monomer 2	500 ns
MD3	Monomer 1	36 ns
	Monomer 2	711 ns
MD4	Monomer 1	400 ns
	Monomer 2	94 ns

B | Fluorescent Behavior of Flugi-2 Molecule

Optimizations in Ground and Excited State

The DFT optimizations and the calculation of the excitation energies was done with the orca programm package[208,209] and the split-valence basis set def2-SV(P). Since, in the orca programm package the range separated functionals for DFT are not implemented I used Gaussian 09[187] DFT (WB97X) and the 6-31G* basis set to optimize the Fluig-2 molecule in ground and excited state.

Additional was used the SORCI (Spectroscopy-Oriented multireference Configuration Interaction) method, as implemented in the ORCA program package. We considered an Complete Active Space with 12 electrons in 12 orbitals, also several roots were tested and the split-valence basis set def2-SV(P) was applied.

For the calculations with implicit solvent the COSMO model was used with an epsilon of 47,8.

In table Table 1 are shown the excitation energies and Stokes Shift of the flugi molecule calculated by several methods based on different geometries. Looking at the excitation energies of the CC2 gas phase geometries, the CC2 and SORCI calculations agree with the experiment (Flugi in Decane). In the CC2 calculations not only the Stokes shift (relative values) is reproduced also the absolute values reproduce the experiment, contrary to the SORCI calculations, here the stokes shift is well reproduced but the excitation energy and fluorescence are shifted by 0.2 eV in comparison to the experiment. DFT with range seperated functional (WB97X) reproduce the Stokes Shift, however the absolute values are shifted by 0.5 eV compared to the experiment. Even the semiempirical method OM2 and DFTB are able to reproduce the Stokes Shift of the experiment. The absolute values of OM2 and DFTB with GGA-functionals are 0.2 - 0.3 eV shifted compared to the experiment. While the Stokes Shift is overestimated and the absolute values of the DFTB with range seperated functionals is shifted about 0.5 eV compared to the experiment. In summary, the CC2 geometries are reasonable and the semiempirical Method OM2 is good enough to calculate the qualitative Absorption and Fluorescence energies of the MD trajectories.

Table 1: Calculations of the excitation and fluorescence energies of flugi-2 with different methods. Geometry optimization was done with WB97X (6-31G*) and CC2 (def2-TZVP). Single point energies based on these optimized structures with different methods.

		excitation energy [eV] (oszillator strength)		
optimization method	method for excitation energy	**Optimized in ground state**	**Optimized in first excited state**	**Stoke shift**
Gasphase CC2	**CC2**	3.22 (0.478)	2.56 (0.361)	0.66
	WB97X	3.88 (0.535)	3.22 (0.416)	0.66
	SORCI root3	3.04 (0.521)	2.36 (0.366)	0.68
	SORCI root5	3.08 (0.588)	2.37 (0.411)	0.71
	CIS	4.17 (0.307)	3.56 (0.537)	0.61
	DFTB *range-seperated*	3.56 (0.714)	2.72 (0.666)	0.84
	DFTB	2.80 (0.327)	2.19 (0.216)	0.61
	OM2	3.57 (0.307)	2.88 (0.248)	0.69
	Experiment (in Decane)	3.26	2.61	0.65
Gasphase WB97X	**WB97X**	4.17 (0.466)	3.35 (0.460)	0.82
	SORCI root5	3.08 (0.3247)	2.50 (0.4659)	0.58
Gasphase B3LYP	**B3LYP**	3.29 (0.459)	-	-
	WB97X	4.003 (0.4971)	-	-

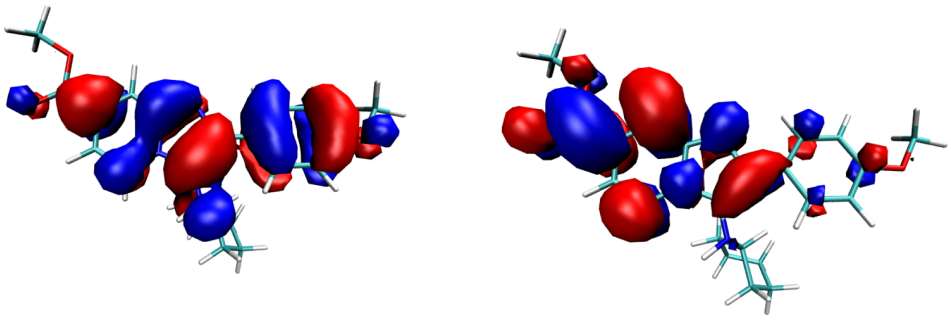
Table 2: Single point energies of CC2-geometries in gas phase. The excitation and fluorescence energies are calculated in gas phase and with implicit solvent DMSO (COSMO) and WB97X-geometries in gasphase and optimized with COSMO

Gasphase CC2	**WB97X-COSMO**	3.76 (0.569)	3.05 (0.438)	0.71
	SORCI-COSMO root5	3.05 (0.591)	2.35 (0.415)	0.70
In DMSO Wb97X-COSMO	**WB97X-COSMO**	3.97 (0.527)	2.93 (0.566)	1.04
	SORCI-COSMO root5	3.33 (0.535)	2.44 (0.528)	0.89
	Experiment (in DMSO)	3.18	2.30	0.88

Determined Parameters of Flugi-2 Force Field

Table 3: ground state parameters for bonds of heavy atoms

bond	force constant [kJ/nm^2/mol]	b0 [Å]
C9-N3	345013	1.3703
C9-N1	361163	1.3854
C8=C9	376476	1.3863
C2-N1	361163	1.3795
N1-C3	446014	1.4460
C1=C2	427856	1.3608
C4-C1	400325	1.3878
C4-C6	306771	1.4503
C5=C4	421747	1.3604
C6-O1	521828	1.2242
C6-O2	328528	1.3554
O2-C7	252295	1.4310
C3-C5	386434	1.3993
C3=N2	412459	1.3379
N2-C8	387104	1.3325
C8-C10	324595	1.4466
C10-C11	376476	1.3990
C14=C10	386434	1.3794
C11=C12	400325	1.3855
C12-C13	386434	1.4040
C13-O3	323088	1.3574
C13=C15	386434	1.3890
C15-C14	386434	1.3944



HOMO LUMO

Figure 1: HOMO and LUMO of the CC2 optimized ground state

Table 4: excited state parameters for bonds of heavy atoms

bond	force constant [kJ/nm^2/mol]	b0 [Å]
C1=C2	400325	1,3758
C4-C1	386434	1,3629
C2-N1	356477	1,3709
N1-C3	361163	1,3900
C9-N1	319490	1,4385
C3-C5	352209	1,4172
C3=N2	375723	1,3563
C5=C4	344511	1,3986
C4-C6	315808	1,4456
C6-O1	510448	1,2293
C6-O2	323088	1,3394
O2-C7	252295	1,4200
N2-C8	375723	1,3482
C8=C9	352209	1,3854
C8-C10	338151	1,4322
C9-N3	393547	1,3289
C10-C11	352209	1,4426
C14=C10	376476	1,4262
C11=C12	400325	1,4173
C12-C13	376476	1,4477
C13=C15	376476	1,4464
C13-O3	323088	1,3400
C15-C14	386434	1,4247

Table 5: C8-C10 dihedral angle parameter ground state

dihedral parameter ground state	phi0	cp	mult
N2-C8-C10-C11	180.000	4.079	2
N2-C8-C10-C14	180.000	4.079	2
C9-C8-C10-C11	180.000	4.079	2
C9-C8-C10-C14	180.000	4.079	2

Table 6: C8-C10 dihedral angle parameter excited state

dihedral parameter excited state	phi0	cp	mult
N2-C8-C10-C11	180.000	6.245	2
N2-C8-C10-C14	180.000	6.245	2
C9-C8-C10-C11	180.000	0.965	2
C9-C8-C10-C14	180.000	0.965	2

Table 7: Fitted charges of non-hydrogen atoms of ground state and excited state. This charges were used for the parametrization of the flugi molecule. The atoms C9, C11, C13, C14, C17, C18, C20 become positively charged, while the atoms C10, C12, C15, C19, C21, C22 become negatively charged. This means every second atoms is negatively charged while the direct neighbor is positively charged. his is reproduced by the HOMO-LUMO plot of the CC2 structures, here, only at every second atom at the benzene ring we have an electron density (see Figure 1).

	ground state (HF)	excited state (DFT)	difference	excited state with COSMO (DFT)
-0,193	-0,006	-0,187	-0,117	
-0,180	-0,440	0,260	-0,310	
0,059	0,150	-0,092	-0,064	
0,469	0,564	-0,095	0,791	
-0,162	-0,545	0,384	-0,438	
-0,261	-0,384	0,122	-0,575	
0,863	1,063	-0,200	1,059	
-0,598	-0,645	0,047	-0,736	
-0,418	-0,318	-0,100	-0,339	
-0,034	-0,258	0,224	-0,269	
-0,624	-0,669	0,045	-0,838	
0,149	0,150	-0,002	0,087	
0,175	0,674	-0,499	0,879	
-0,850	-1,235	0,385	-1,328	
0,000	-0,797	0,797	-0,662	
-0,061	0,616	-0,677	0,443	
-0,398	-1,311	0,913	-1,244	
0,503	1,248	-0,745	1,254	
-0,061	0,941	-1,002	0,828	
-0,398	-1,100	0,702	-1,111	
0,430	0,107	0,323	0,199	
0,386	0,515	-0,129	0,563	
-0,083	1,131	-1,214	1,223	
-0,083	0,685	-0,769	0,773	
-0,033	-0,502	0,469	-0,478	
-0,014	0,754	-0,767	0,787	
-0,033	-0,637	0,604	-0,633	
-0,411	-0,290	-0,121	-0,343	
0,078	-0,620	0,698	-0,612	

Bond Lengths and Dihedral Angles of Ground and Excited State

Table 8: Bond length of CC2 structure, minimized MM structure and averaged structure of MM-MD in ground state in gas phase [Å]

bond	CC2	MM minimized	MM-MD
C9-N3	1,3915	1,3886	1,3875
C9-N1	1,3741	1,3641	1,3665
C8=C9	1,4065	1,4168	1,4169
C2-N1	1,3739	1,3731	1,3730
N1-C3	1,4193	1,3908	1,3921
C1=C2	1,3676	1,3705	1,3723
C4-C1	1,4194	1,4194	1,4289
C4-C6	1,4752	1,4835	1,4786
C5=C4	1,3904	1,3990	1,3977
C6-O1	1,2224	1,2268	1,2285
C6-O2	1,3576	1,3723	1,3716
O2-C7	1,4406	1,4452	1,4458
C3-C5	1,3940	1,3981	1,3998
C3=N2	1,3388	1,3341	1,3387
N2-C8	1,3670	1,3744	1,3696
C8-C10	1,4572	1,4555	1,4607
C10-C11	1,4075	1,4107	1,4112
C14=C10	1,3997	1,3918	1,3916
C11=C12	1,3872	1,3953	1,39789
C12-C13	1,4014	1,4047	1,4076
C13-O3	1,3642	1,3640	1,3627
C13=C15	1,3999	1,3988	1,3924
C15-C14	1,3947	1,3984	1,4041

List of Abbreviations

MD-Simulation Molecular Dynamic Simulation

GROMACS Groningen Machine for Chemical Simulations

TIP3P Watermodel

DFT Density Functional Theory

LDA Local Density Approximation

GGA Generalized Gradient Approximation

SCF Self-Consistent Field

DFTB Density Functional Tight Binding Theory

CI Configuration Interaction

CASSCF Complete Active Space Self-Consistent Field

MRCI Multi-Reference Configuration Interaction

OM2/MRCI Orthogonalization Model 2/Multi-Reference Configuration Interaction

SORCI Spectroscopy-Oriented Configuration Interactions

MP2 Møller-Plesset Perturbation

QM/MM Quantum Mechanic/Molecular Mechanic

RSBH Protonated Schiff base

BLA Bond Length Alternation

BR Bacteriorhodopsin

ChR-2 Channelrhodopsin-2

HKR Histidin Kinase Rhodopsin

RR Response Regulator

DAG diacyglycerol

cGMP cyclic Guanosine Monophosphate

DB Double bond

SB Single bond

Acknowledgements

Firstly, I would like to express my sincere gratitude to my advisor Prof. Dr. Marcus Elstner, for the continuous support of my Ph.D study and related research, for his patience, motivation, and scientific advise. His guidance helped me in all the time of research over the last years. He gave me the freedom to expand my knowledge based on my own interests.

Beside my advisor, I would specially like to thank Dr. Igor Schapiro, who provided me an opportunity to join his team at the Fritz-Haber Institute (Hebrew University) in Jerusalem. During my stay, I felt very comfortable and welcome in the group and enjoyed the discussions during the group meetings. I gained a lot experience about theoretical methods and learned a lot about the Jewish people, and I had the chance to find new friends. I would also like to thank Dr. Veniamin Borin, who helped me a lot with the technical details and the fruitful discussions during my stay.

My sincere thanks also goes to my colleagues in Karlsruhe for the nice atmosphere, their insightful comments and encouragement.

Special thanks goes to Beatrix Bold, Mila Andreeva and Violetta Schneider, from our "Girls Tower", who have been very busy proofreaders. I am grateful for their inspiring guidance, constructive criticism and friendly advice during the thesis and for sure, for all the fun we have had in the last years.

I would also like to thank Sabine Holthoff, who had always an open ear and supported me with all kinds of administrative affairs.

I thank my collaborators Prof. Dr. Tomasz Wesolowski and Emilie Chalaye from the University in Geneva for their contribution to our good work.

I am thankful for the Minerva foundation and the RTG 2039 for their financial support during my abroad in Jerusalem.

My special thanks goes to my girls Andrea Lauer, Bettina Olshausen and Carolin Heiler for their mental support and the wonderful hours, which we have been able to spend together.

I am grateful for the support, patience and motivating words of my family.

At the end, I want to thank my husband for all the nerdy discussions in the last years, the hours spending back to back in our small office at our home and for the mental support, helping me to believe in myself and my opportunities. I would not like to miss any of these moments.