QM/MM Simulations of Channelrhodopsins – Elucidating Structure and Spectroscopic Properties

Zur Erlangung des akademischen Grades eines
DOKTORS DER NATURWISSENSCHAFTEN

(Dr. rer. nat.)

Fakultät für Chemie und Biowissenschaften
Karlsruher Institut für Technologie (KIT) – Universitätsbereich
genehmigte

DISSERTATION

von
Kai Welke
aus
Haldensleben

Dekan:	Prof. Dr. Peter Roesky
Referent:	Prof. Dr. Marcus Elstner
Korreferent:	Prof. Dr. Willem Klopper
Tag der mündlichen Prüfung:	20. Dezember 2013

Zusammenfassung

In der vorliegenden Arbeit werden Strukturen mikrobieller Rhodopsine modelliert, um den Zusammenhang zwischen strukturellen und spektroskopischen Eigenschaften besser zu verstehen. Hierzu werden Multiskalenmethoden benötigt, die klassische und quantenmechanische Ansätze kombinieren (QM/MM). Dazu generieren wir Strukturmodelle, optimieren deren Strukturen und berechnen ihre UV/vis und Schwingungspektren, um sie mit experimentellen Beobachtungen zu vergleichen.

Bacteriorhodopsin ist das am besten untersuchte Rhodopsin und eignet sich daher vortrefflich als Referenz für die Untersuchung anderer, noch wenig untersuchte Proteine, zum Beispiel der Channelrhodopsine. Channelrhodopsine sind direkt licht-gesteuerte Ionenkanäle, die die Orientierung von einzelligen Algen in der Natur steuern. Natürlich vorkommende Channelrhodopsine funktionieren am besten unter blauem Licht. In den Neurowissenschaften werden Channelrhodopsine für die Anregung spezifischer Zelltypen mit Licht bereits erfolgreich angewendet, wofür jedoch Varianten geeigneter sind, die sich mit rotem Licht anregen lassen. Dabei ist bisher kaum verstanden, welche Faktoren die Absorptionsspektren der Channelrhodopsine beeinflussen (*Color Tuning*). Vergangene Studien haben gezeigt, wie wichtig die Bindungstasche eines Rhodopsins für das Color Tuning ist. Dabei war die Struktur der Bindungstasche der Channelrhodopsine zunächst noch unbekannt, und ist auch heute noch sehr wenig erforscht. Ausgehend von der bekannten Struktur des Bacteriorhodopsins modellieren wir die Bindungstaschen der Channelrhodopsine durch Mutationen. Wir zeigen, dass sich durch die Bindungstaschenmodelle das Color Tuning in den Channelrhodopsinen simulieren lässt und zeigen die Grenzen eines mutationsbasierten Ansatzes für die Optimierung der Channelrhodopsine auf.

Die QM/MM Methoden, die im Verlaufe der Dissertation angewendet werden, eignen sich hervorragend, um die experimentellen Beobachtungen mit einem Strukturmodell zu verknüpfen und dabei neue Sichtweisen zu eröffnen. Wir zeigen, dass sich die Methode DFTB3 dazu eignet, Schwingungsspektren von Carbonsäuren zu berechnen und die Einflüsse von Wasserstoffbrückenbindungen auf die Schwingungsspektren akkurat zu beschreiben. Zunächst zeigen wir am Beispiel von Bacteriorhodopsin, wie sich kleine strukturelle Änderungen im Wasserstoffbrückennetzwerk auf Spektrum auswirken. Das DC-Gate ist ein entscheidendes strukturelles Motiv für die Kinetik des Ionenkanals in den Channelrhodopsinen. Angespornt durch Widersprüche des bisherigen Modells zur Erklärung der Schwingungsspektren des DC-Gates mit neueren Erkenntnissen, schlagen wir eine Alternative vor, welches mit neuesten Simulationen und Experimenten gut übereinstimmt und in der Lage ist, die Schwingungsspektren zu erklären.

Obwohl Bacteriorhodopsin so gut verstanden ist, gibt es immer noch offene Fragen, gerade in Hinblick auf die Schwingungsspektren. Der Ursprung einer breiten und schwachen Bande im Schwingungspektrum ist bisher noch ungeklärt, wobei aufgrund der Eigenschaften dieser Bande die Vermutung nahe liegt, dass die N-H Streckschwingung der Schiff'schen Base, durch welche das Chromophor an das Apoprotein gebunden ist, diese Bande verursacht. Prinzipiell eignen sich die von uns verwendeten QM/MM Methoden gut, die N-H Streckschwingung zu beschreiben, gelangen hier aber an ihre (vorläufige) Grenzen, so dass auch unsere Simulationen den Ursprung dieser Bande nicht aufklären können

Abstract

The work described in the present thesis considers structural modelling of microbial rhodopsins, to better understand the relation between structural and spectroscopic properties. To achieve that, multi-scale approaches are required, that combine both classical and quantum-chemical ideas (QM/MM). Therefore, we build structural models, optimize their structure, and calculate their UV/vis and vibrational spectra to compare to experimental observations.

Bacteriorhodopsin is the rhodopsin, which is understood best. Therefore, it serves as an excellent reference for the studies of the channelrhodopsins, which are not so well understood. Channelrhodopsins are light-gated ion channels, which control the motion of single-cellular green algae. Naturally-occurring channelrhodopsins are most active under blue light. In neuroscientific experiments, the channelrhodopsins are used successfully to excite specific cell types via light. However, better suited for these experiments are variants, that are excitable by red light. Up to now, the factors, which influence the absorption spectra of the channelrhodopsins are poorly understood (*color tuning*). Previous studies showed, how important the binding pockets of the rhodopsins are for color tuning. Because the structure of the binding pockets of the channelrhodopsins was previously unknown and is still not well understood, we built structural models by mutations, starting from the known structure of bacteriorhodopsin. We show, that by using binding pocket models, we can simulate the color tuning in the channelrhodopsins, while showing the limits of an approach based on mutations for engineering optimized channelrhodopsins.

The QM/MM methods applied throughout this thesis are very suitable to connect experimental observations with underlying structural models and to obtain new insights. We show, that the method DFTB3 is capable to describe the C=O stretch vibration in carboxylic acids and effects of hydrogen bonds on the vibrational spectra. We show, how small structural changes in the structure of the hydrogen-bonded network affect the vibrational spectrum. In the channelrhodopsins, the DC gate is an important structural motif for the ion channel's kinetics. Previous models for the explanation of vibrational spectra are inconsistent with newer insights obtained from simulations and experiments. Inspired by these contradictions, we propose an alternative model for the DC gate, which overcomes all shortcomings of the previous model.

Although bacteriorhodopsin is well understood, there are still open questions, especially considering vibrational spectroscopy. The origin of a broad and weak band appearing in the spectrum is still unknown. Its properties suggest that the N-H vibration of the Schiff base, by which the chromophore is bonded to the apoprotein, is involved. Our QM/MM methods are generally able to describe the N-H stretch vibration of the Schiff base but reach their (preliminary) limits during this study. Therefore, our simulations cannot explain the occurrence of the band either.

Contents

IV Appendix

Part I

Introduction

CHAPTER 1

Microbial Retinylidene Proteins

Over three hundred retinylidene proteins can be found in every domain of life, also known under the synonymous names: rhodopsins, retinal proteins, or opsins (which strictly speaking only refer to the apoprotein).[1] The DNA sequences of over 1000 opsins are known.[2] All have in common, that they consist of a transmembrane region of seven αhelices, which form an internal pocket, in which a retinal is covalently bound to a lysine residue within the seventh transmembrane helix, via a carbon-nitrogen double bond, a motif to which the chemist refers to as *Schiff base* or *imine* (Fig. 1.1(c)). The photoisomerization of the retinal and the subsequent reaction of the apoprotein is the key to the function of all known retinylidene proteins.

While in this work, the channelrhodopsins are the focus of attention, a general introduction into rhodopsins and a more detailed introduction into bacteriorhodopsin is not wasted here. The vast amount and detail of knowledge about the structure and function of bacteriorhodopsin makes it an excellent model system for both experimental and computational studies. Therefore it is often used as reference in the study of newly discovered rhodopsins – a common practice, which is also exercized in this work.

1.1 Rhodopsins

The rhodopsins are divided into two families, based on their primary sequence. Type 1 (microbial) rhodopsins are found in archaea, bacteria and fungi. They assume diverse roles, such as passive cation channels, active proton, cation or chloride pumps and photoreceptors.[1] So far, all type 2 rhodopsins have been found in higher eukaryotes and are photoreceptor proteins, embedded in the retina in the eyes, in the brains of mice and men or other tissue of lower vertebrates. Furthermore, all type 2 rhodopsins are coupled to a member of the family of *guanoside nucleotide-binding proteins* (G-proteins), which transmit signals from the outside to the inside of a cell. Thus, type 2 rhodopsins also belong to the family of G-protein coupled receptors (GPCRs). While type 1 and type 2 rhodopsins both consist of seven αhelices, the tertiary structures, i.e. the way these helices are arranged, are different. Another difference among the two families is rooted in the configuration of the chromophore. In most type 1 rhodopsins, the active conformation of the retinal is all-*trans* (Fig. 1.1(b)). After photoexcitation, the retinal is isomerized mainly to 13-*cis*. The retinal re-isomerizes to all-*trans* while staying attached to

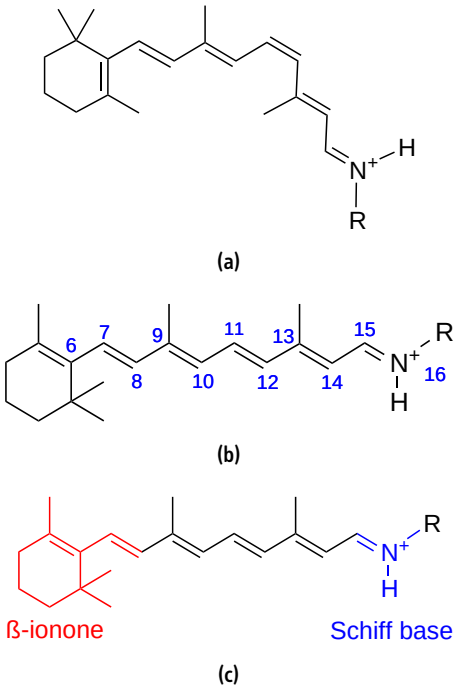


Figure 1.1 The retinal cofactor of the rhodopsins in its most prevalent configurations: **(a)** the 11-*cis* configuration found in the ground state of type 1 rhodopsins and **(b)** The all-*trans* configuration as found in the ground states of type 2 rhodopsins. The heavy atoms of the retinal are usually numbered as shown. **(c)** The different regions of the retinal which are especially important for its properties.

the opsin. Thus, the whole process of excitation, isomerization and re-isomerization and thus, the coupled protein function can be repeated in a cyclic fashion (photocycle). The retinal in type 2 rhodopsins isomerizes from 11-*cis* to all-*trans* and is hydrolyzed from the opsin in the last part of the functional sequence. After that, it has to be re-isomerized by several different enzymes in a different part of the organism before being re-attached to the opsin.[3]

So far, no DNA sequence has been found that would match to both, type 1 and type 2 rhodopsins. Consequently, it is a matter of debate whether both types belong to a superfamily of proteins, sharing the same evolutionary precursor (divergent evolution). Over time, sequence homology between the two types decreased to an extent, that it is practically non-existent today. The alternative scenario is convergent evolution, which means, both families evolved independently, while the chromophore and transmembrane property act as restraints on the overall structure.[4,5] In other words, in order to obtain functional photoactivatable proteins with the retinal as chromophore, they have to surround the cofactor in the seven transmembrane helix fashion of the rhodopsins.

1.2 Bacteriorhodopsin

The most studied and best understood microbial rhodopsin is the bacteriorhodopsin (BR). It was first found in *Halobacterium halobium* in the early 1970s.[6] Despite the name, halobacteria belong to the

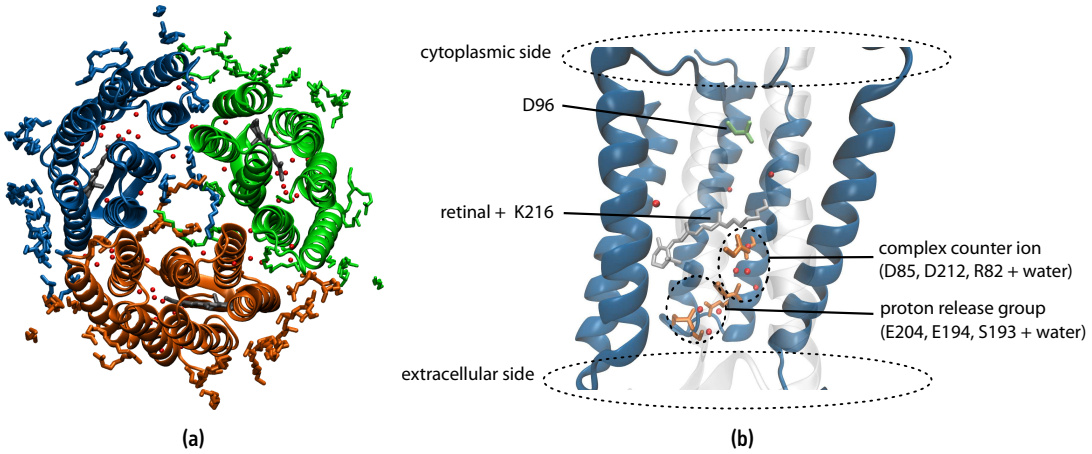


Figure 1.2 The three-dimensional structure of bacteriorhodopsin, as determined by X-ray crystallography.[14] **(a)** the trimeric unit as embedded in the purple membrane, **(b)** functional important parts of the monomer.

domain of Archaea rather than Bacteria.[a] Halobacteria are extremophiles, that prosper under high salt concentrations, e.g. in the Great Salt Lake in Utah, USA or in the Dead Sea. The reddish purple color of large patches of halobacteria emerges from the bacteriorhodopsin pigment, which has an absorption maximum near 570 nm. The protein is arranged in trimeric units (Fig. 1.2(a)) in the so-called purple membrane in a two-dimensional hexagonal lattice in a 3:1 ratio of protein to lipid (by weight).[6,9–11] To provide energy for metabolism, motion, and reproduction organisms utilize the adenosine triphosphate (ATP) molecule, which is often termed the energy "currency" of life. The synthesis of ATP by ATP synthase requires energy, which is provided in the form of an electrochemical proton gradient. In all photoautotrophic organisms, including most plants, algae, and cyanobacteria, this electrochemical proton gradient is generated in the chloroplasts and mitochondria during photosynthesis and cellular respiration.

At high salt concentration, the conditions become too harsh for photoautotrophs, but ideal for the extremophile haloarchaea. As an alternative source of a proton gradient, BR uses the energy of light to actively transport protons from the cytoplasm to the extracellular side against a pH gradient.[12] The first low-resolution, electron density map of BR and the purple membrane appeared 1975, making BR the first membrane protein, for which any structural information was available.[13] Thanks to the highly organized purple membrane, BR is one of the few membrane proteins, for which X-ray structures are available, with high resolutions of up to 1.55 Å.[14]

BR is a relatively small protein (24 kDa),[b] containing 248 amino acids arranged in seven αhelices that span across the purple membrane. Several strongly hydrogen-bonded internal water molecules are resolved in X-ray structures and play important roles in the proton transfer across the protein. The αhelices are not perfectly straight. In helices 2, 3 and 6, a proline residue is the origin of kinks, whereas in helix 7, at the position of the retinal binding Lys 216, a single turn of a π-helix breaks the αhelical

[a] In the past, archaea and bacteria were fused under the name *archaebacteria*, but advancements in phylogenetic research showed, that bacteria and archaea have evolved distinctly and should therefore be regarded separately.[7,8]

[b] 1 Dalton is equivalent to the atomic mass unit definition, i.e. 1/12 of the mass of 1 carbon-12 atom.

structure.[15] The retinal binding site also separates the protein into an extracellular and a cytoplasmic part (Fig. 1.2(b)).

1.3 The Photocycle of Bacteriorhodopsin

When the retinal chromophore is electronically excited by light, it isomerizes and the whole protein undergoes specific structural changes that accompany the transport of a proton from the cytoplasmic to the extracellular side of the membrane. Several structural intermediates have been trapped at low temperatures and subsequently characterized by X-ray crystallography, nuclear magnetic resonance (NMR), Fourier-transformed infrared (FTIR), or Raman spectroscopy. While there is no universal definition of the photocycle, since the detection and characterization is strongly dependent on experimental conditions, a representation for a widely accepted model is shown in Figure 1.3.[15,16] It should be stressed again, that the following discussion restricts itself to the major aspects of the photocycle, while each transition and intermediate possibly involves complications, spectroscopically silent intermediates, multiple conformations, etc. For a more thorough discussion, the reader is referred to refs. 17 and 15.

When a photon of sufficient wavelength (λ_{max} = 568 nm = 2.18 eV $\approx$ 50 kcal/mol) is absorbed by the retinal chromophore, it becomes excited from the electronic ground state (S_0) to the first excited state (S_1). This change is accompanied by a shift of electron density in the polyene chain and a change in bond order.[18] Sometimes, the label H, I, and J are applied to proposed excited states (rather than intermediates) between the dark state and the K intermediate, but due to the short transition time between the dark state and K, their spectroscopic characterization has been all but robust. The isomerization of the C13=C14 double bond occurs within 3 ps in the relaxation of the putative J state to the K intermediate, which is the first cryotrappable intermediate.[19] The retinal adopts a twisted 13-*cis*,15-*anti* conformation. Structural differences between K and the dark state are minor and restricted to the Schiff base.[20]

The excess energy from the photon is stored in the K intermediate and is used for all subsequent steps of the photocycle. The next step, the K→L transition, possibly involves the relaxation of the chromophore to a more planar 13-*cis* conformation in preparation of the first proton transfer step. The nature of this transition is complex and highly dependent on experimental conditions, so that details are still debated.[21–23] The L→M transition occurs on a 10 µs timescale and involves the transfer of a proton from the retinal to Asp 85, which acts as proton acceptor.[24–28] At the same time, a proton is released to the extracellular side of the membrane from the aptly named proton release group (PRG), that involves residues Glu 194, Glu 204, Ser 193 and one or more water molecules.[29–31] During the M→N transition, the retinal gets reprotonated by Asp 96 on a millisecond timescale.[32,33] This transfer step occurs over a large distance (12 Å) and possibly involves a transient water chain.[34] Asp 96 is then protonated again from the cytoplasm. During the N→O transition, the retinal chromophore reisomerizes into the all-*trans* conformation, possibly in a twisted conformation.[35] In the last step of the photocycle, the O→BR transition, Asp 85 looses its proton to the PRG which completes the cycle. However, compared to the first half of the photocycle, information about the last step of the photocycle is rather sparse.[15]

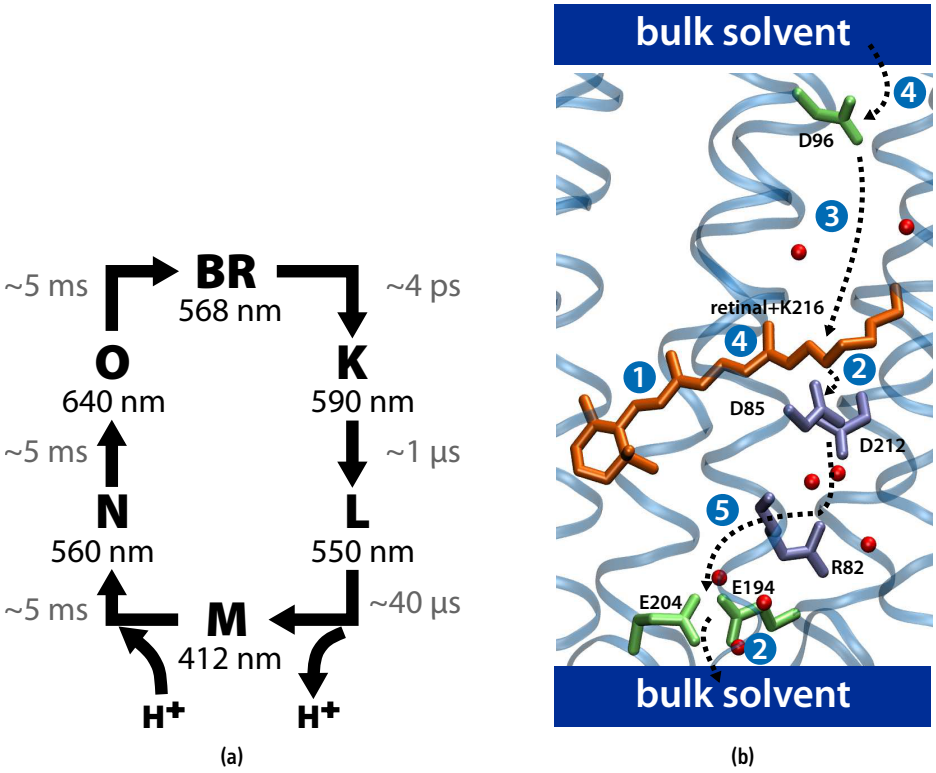


Figure 1.3 Simplified model of the photocycle of BR. **(a)** schematic representation with names of important intermediates, their absorption maxima and approximate lifetimes at room temperature. Possible backtransitions between intermediates are omitted. **(b)** proton pathway through BR. (1) The first step of the photocycle after photoexcitation of the chromophore is the isomerization of all-*trans* retinal to 13-*cis* in the BR→K transition. (2) In the L→M transition, one proton is transferred from the retinal to Asp 85. At the same time, one proton is released from the PRG, which carries one positive charge and includes Glu 204, Glu 194 and water molecules. (3) In the M→N transition, the Schiff base gets reprotonated by Asp 96. (4) In the N→O, the retinal re-isomerizes to all-*trans* and Asp 96 picks up one proton from the cytoplasm. (5) In the last step of the photocycle, one proton is transferred from Asp 85 to the PRG via Arg 82 and possibly one or more water molecules.

CHAPTER 2

Channelrhodopsins – Light-gated Ion Channels

The channelrhodopsins (ChRs) are also microbial rhodopsins, but are evolution-wise rather distant cousins of BR with unique features and applications, that allow them to occupy an introductory chapter on their own. ChRs have gathered a lot of attention from scientists of many fields over the last decade. What makes the ChRs attractive to study is the fact, that there is an application for them with a clearly formulated goal in neuroscientific and medical experiments (*optogenetics*).

The understanding of the relationship between the function and characteristics of the ChRs and the underlying structural features is very lacking. While the first application was already reported in 2005, three years after the initial description of a channelrhodopsin,[36] the first useful structural models appeared another seven years later in 2012.[37,38] On one hand, much information is available regarding the electrochemical properties of the ChRs, such as photocurrents, kinetics, ion selectivity, expression levels and optimal excitation wavelength – optogeneticists gained much experience in effectively manipulating theses properties to their advantage. On the other hand, very little is known about why certain changes in the amino acid sequence cause certain changes in properties. The desire to manipulate and engineer new optogenetic tools entails the desire to understand the mechanism of the ion channeling in the ChRs. Due to their high resolution, computer simulations are the perfect tools to provide such a fundamental understanding in close collaboration with experiments.

The study of the channelrhodopsins involve scientists from many fields, ranging from computational chemistry to neuroscience and psychiatry. What is accepted as common language in one field may be completely incomprehensible to another. Therefore, it is the aim of this chapter to introduce the most important aspects of the ChRs in a basic fashion. At the end of the chapter, the open issues revolving around the ChRs are emphasized.

2.1 Microbial Rhodopsins as Cation Channels in *Chlamydomonas reinhardtii*

Reading early literature on rhodopsins in the green alga *Chlamydomonas* can be confusing, since it was unclear for a long time, what type of rhodopsin was responsible for controlling the movement of the organism. Additionally, multiple entries in gene databases referring to the same protein complicate research. Although this topic is not of direct concern, this section helps to establish nomenclature of

the ChRs as well as help to put them into relationship to other rhodopsins. Another retrospective on the events that lead to the discovery of the ChRs can be found in ref. 39.

The eucaryotic, unicellular green-alga *Chlamydomonas reinhardtii* is intensively studied in biology since its first description in 1888.[40] Nowadays it is used as model organism to address issues regarding motility, reproduction, and response of organisms to light. *C. reinhardtii* is photomobile, i.e. it orients towards or away from a light source (positive and negative phototaxis), and avoids too bright areas (photophobic behavior). In the membranes of the eyespot of *C. reinhardtii*, a red-colored membrane stack filled with carotenoids is embedded, containing an estimated amount of 10^5 receptor pigments.[41] The eyespot acts as a quarter-wave plate to modulate the light onto the receptor pigments. First hints of a rhodopsin in *C. reinhardtii* were recorded in 1984, when the behavioral response of blind mutants of *C. reinhardtii* could be restored by adding retinal. The action spectra, i.e. the frequency dependent response of the organism to light, looked nearly identical for the variants with 11-*cis* retinal and all-*trans* retinal. The opsin shift, the difference between the absorption maximum of the chromophore in solution and in the protein environment, was more similar to type 2 (animal) rhodopsins, than it was to BR, a type 1 rhodopsin. This lead to the hypothesis, that the rhodopsin in *C. reinhardtii* is an animal-type rhodopsin. The first sequence of a rhodopsin-like protein was deposited under the name *chlamyopsin*, encoding two proteins with alternative splicing: chlamyopsin-1 (cop1) and chlamyopsin-2 (cop2). At least for cop2, it was established by gene silencing, that it is involved in the assembly of the photosystem I complex of *C. reinhardtii*, but is not crucial.[42] Cop1 and cop2 are very distinctive from other known rhodopsins.[43] Few similarities were found between chlamyopsin and type 1 rhodopsin sequences and only four transmembrane helical regions could be located. Overall, the organization of the gene resembles invertebrate (type 2) rhodopsins, rather than microbial rhodopsins, e.g. the G-protein binding domain, that couples the signal transducer protein to the rhodopsin in animal-type rhodopsins is conserved in cop1.[44,45] Silencing the chlamyopsin gene did not reduce the electrical response to a light flash (photocurrent) and thus made clear, that this is not the photoreceptor for phototaxis in *C. reinhardtii*.[46]

In 2000, at the Kazusa DNA research institute in Japan, desoxyribonucleic acid (DNA) fragments of *C. reinhardtii* could be sequenced and lay the foundation for the subsequent findings of the ChRs.[47] In the group of Peter Hegemann, two rhodopsin-type DNA sequences were found. The gene encoded by the first sequence was named by them *channelopsin-1* (Chop1).[48,a] Chop1 could be successfully expressed in oocytes of the frog *Xaenopus laevis* in the presence of all-*trans* retinal, reconstituting the protein channelrhodopsin-1 (ChR1).[49] In voltage-clamp experiments, inward currents could be measured, and it was established, that the photocurrent was carried by passive transport rather than active pumping. The unique combination of a photoreceptor and an ion channel has not been found before in any protein. The second sequence, named channelopsin-2 (Chop2)[b] was successfully expressed as channelrhodopsin-2 (ChR2) in *X. laevis* oocytes, and mammalian cells, already suggesting its future application in optogenetic experiments.[50]

Parallel to and separate from the work of Hegemann et al., Sineshchekov and coworkers used the same DNA database[47] as well as their own genomic DNA to identify the two rhodopsins as *Chlamydomonas sensory rhodopsin A* and *B*.[51] Although the names never stuck in the community, it was the

[a] Chop1 was originally called chlamyopsin-3 (cop3), before it was known, that it was an ion channel.
[b] Chop2 was originally called chlamyopsin-4 (cop4).

Table 2.1 Opsin sequences found so far in *C. reinhardtii*, their abbreviations, and reports in the literature. Further rhodopsin sequences and accession numbers for the NCBI GenBank[55] can be found in Table A.1 (p 99).

name(s)	abbreviation	ref.
chlamyopsin, chlamyopsin-1*	cop, cop1	43
chlamyopsin-2*	cop2	45
channelrhodopsin-1, chlamyopsin-3,	ChR1,cop3	48,49
Chlamydomonas sensory opsin A,	CSOA	51
archaeal-type Chlamydomonas opsin 1	Acop1	56
channelrhodopsin-2, chlamyopsin-4,	ChR2	48,50
Chlamydomonas sensory opsin B,	CSOB	51
archaeal-type Chlamydomonas opsin 2	Acop2	56
histidine kinase rhodopsin†,chlamyopsin-5	hkr1,cop5	52,54
chlamyopsin-6	cop6	52,53
chlamyopsin-7	cop7	52

* cop1 and cop2 originate from the same gene and are alternative splicing products.

† cop5 was originally deposited in 2004 and was later determined to be the HKR. The name cop5 was assigned anew in 2007, under the accession number EDP06598. Its sequence is 98 % homologous to HKR.

first proof, that ChR1 and ChR2 were indeed responsible for controlling the photomotility in *C. reinhardtii*. In the first homology modelling study, which the authors combined with DNA sequencing from the same DNA library,[47] Suzuki et al. decided to name the genes *archaeal-type Chlamydomonas opsin 1 and 2* (Acop1 and 2). Again, these names were not used for the ChRs henceforth.

Three additional rhodopsin-like sequences were reported and named chlamyopsin-5,-6 and -7 (cop5 to 7).[52,53] The sequence, which was deposited under the name cop5, was later determined to result from a histidine kinase rhodopsin.[54] Resulting from the sequencing of the complete genome of *C. reinhardtii*, the name chlamyopsin-5 was re-assigned,[53] but may be identical to the histidine kinase rhodopsin, since it shares 98 % sequence homology. Cop6 is another large protein, similar to cop5, and presumably is a histidine kinase rhodopsin as well.[53] The sequence of cop7 is not deposited in any database, and the protein is only mentioned once.[52] It is supposed to have partial sequence homology with cop5 and the same domain organziation. The known rhodopsins in *C. reinhardtii* are summarized in Table 2.1. Their evolutionary distance to other microbial rhodopsins (Fig. 2.1), such as BR, makes it difficult to predict their characteristics based on structural analogies. The protein, which is closest to the ChR in the phylogenetic tree *and* for which a X-ray structure is available, is *Anabaena* sensory rhodopsin (ASR). In fact, using ASR instead of BR as template structure was a key for the successful generation of the first realistic computational model for ChR2.[37]

Most of the research focusses on channelrhodopsins from *C. reinhardtii*, because of the predominant application of ChR2 in optogenetic experiments and because it is the only ChR, for which a three-dimensional structure is available.[38] However, in recent years, channelrhodopsins have been found in other species of the *Chlamydomonas* genus, namely *C. augustae*, *C. yellowstonensis* and *C. raudensis*.[59] Channelrhodopsin-like sequences have also been found in other green algae.[60] For most of

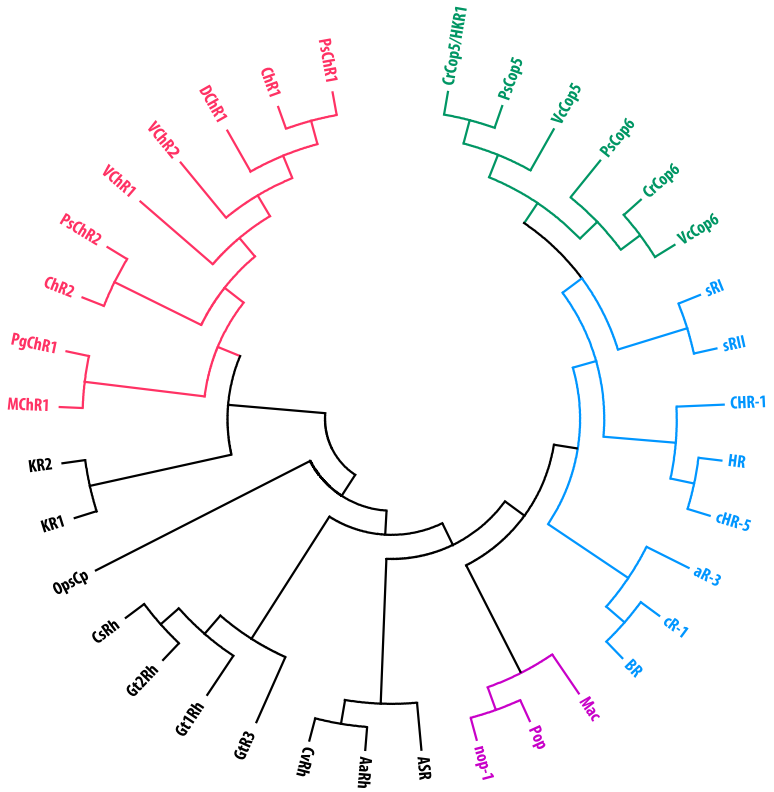


Figure 2.1 Phylogenetic tree of selected microbial rhodopsins. Abbreviations as in Tables 2.1 and A.1. The evolutionary distance between two proteins increase with the number of branches separating them. The algal channelrhodopsins (red) are evolution-wise only distantly related to the archaeal rhodopsins found in halobacteria (blue), including BR. Other rhodopsins, found in *C. reinhardtii* and related organisms, belong to the group of histidine kinase receptors (green). Only a few fungal rhodopsins have been described yet (purple). The relatively new cation pumps KR1 and KR2 are closely related to the ChRs. Multiple Sequence Alignment has been performed with ClustalW.[57] The phylogenetic tree has been drawn using the neighbor-joining method in the MEGA software package.[58]

them, only their DNA sequence is known, except the channelrhodopsin found in the freshwater green alga *Mesostigma viride*, which is also characterized spectroscopically.[61] Volvox channelrhodopsin-1 (VChR1) gathered some attraction due to its application in optogenetic experiments.[62]

2.2 Properties of *C. reinhardtii* Channelrhodopsins

In the following section, the known properties in terms of structure, spectroscopy and conducting properties are recapitulated. Unfortunately, the current state of knowledge is quite diffuse, since many experiments are done not on the ChRs themselves, but on mutants or chimeras, so that it is not always straightforwardly possible to translate the conclusions of these experiments to the original proteins. Furthermore, experiments are done either with purified proteins (spectroscopy) or expressed in a variety of host cells. It may very well be possible, that this different environment causes the ChRs to respond differently than they would in *Chlamydomonas*. However, since optogenetic experiments consists precisely of a heterogeneous expression of ChRs, these information are still very valuable.

2.2.1 Structure

Channelrhodopsin-1 and -2 are directly light-gated ion channels, a combination, which is unique in nature. ChR1 is a 76.4 kDa protein of 712 amino acids length, of which residues 76 to 309 (one third of the amino acids) comprise the seven transmembrane helix.[49] Compared to other microbial rhodopsins, like BR, it has a long C-terminus of yet unknown function. Truncated ChR1, with only the first 346 amino acids, generate photocurrents just as well. In the N-terminal region, an unusual number of glutamate residues can be found, which makes an assignment of the transmembrane regions difficult. It shares a 25 % sequence identity and 47 % similarity with BR. Some amino acids, that interact with the retinal are conserved, such as the retinal binding site, the Lys 296 (Lys 216 in BR), the counterions Glu 162 and Asp 292 (Asp 85 and 212 in BR) and Arg 159 (Arg 82 in BR). Other residues, that are important for the function as a proton-pump in BR are exchanged in ChR1, such as His 173 (Asp 96 in BR), because they are not required in a channel.[49] Channelrhodopsin-2 is slightly larger than ChR1 (737 amino acids, 77 kDa). Overall 55 % of its amino acids are identical with ChR1, only 22 % with BR (40 % similarity). Again, truncation to the first 315 amino acids did not affect the photocurrents negatively.

The two-dimensional electron density map for ChR2[63] and the X-ray structure for the ChR1/ChR2 hybrid (first 5 helices of ChR1, last 2 helices of ChR2)[38] indicate a dimeric unit with helices 3 and 4 being the interface for the protein – protein binding (Fig. 2.2(a)), notably via highly conserved cysteine residues 73 and 75 (ChR1 numbering). Calculations of the electrostatic surface potential revealed an electronegative pore, involving helices 1, 2, 3 and 7, which contain several glutamate residues, that are aligned in a pseudo-one dimensional way (Fig. 2.2(b)).[37,38]

The protonation states of titrable residue is an important structural property, which has a great influence on computer models and therefore also on the interpretation of experiments. Currently, only the protonation state of Asp 192 (Asp 115 in BR) is widely accepted as protonated for both ChR1 and ChR2.[38,64] For the remaining residues, their protonation state belong to the open issues, which are discussed below.

2.2.2 Properties

Due to their suitability for optogenetic experiments, ChR2 is the best examined ChR. Both ChR1 and ChR2 are mostly proton channels,[49,50,66,67] but are able to conduct the physiologically relevant cations as well, albeit with a lower conductance.[66,67] A single ChR2 channel has an estimated conductance of 50 – 250 fS,[50,68–70] which is lower than common ion channels.[71] All ChRs are cation channels, non-selective towards H^+, Na^+, K^+ and Ca^{2+}. Without any membrane potential, ChR2 also pumps protons outwardly,[70] but the mechanism seems to be different from BR as indicated by mutation studies.[72] Under continuous illumination with high intensity light, the response of the ChR decreases from a peak current to a steady state. The ratio between the two currents can be regarded as measure for desensitization,[50] which is the main drawback of ChR2 as optogenetic tool, since the desensitization can be up to 80 %.[50,73] and recovers after 25 s in the dark.[69] The depolarization of membranes is not only controlled by single channel conductance, but also by expression levels in membranes, which are different for each ChR. ChR2 trafficks well into the membrane, which made it the optogenetic tool of choice in the past. The ion channel opens on a time scale of 200 µs and relaxes with a time-scale of 20 ms.[68] The kinetics of the channel can be easily influenced by single mutations. Mutating

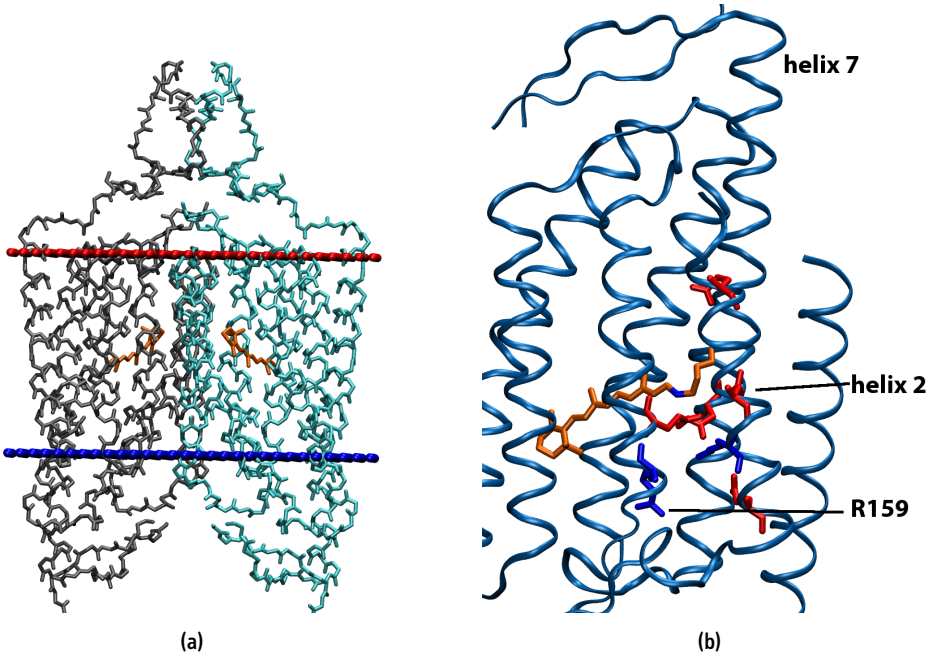


Figure 2.2 **(a)** Embedding of the ChR1/ChR2 chimera as dimeric unit in a membrane. The long loop between helix 2 and 3 protrudes into the extracellular side (top). The cytoplasmic side of the membrane is at the bottom. Coordinates downloaded from the OPM database (orientation of proteins in membranes).[65] **(b)** Highlighted unique features of ChR, including the unusual number of charged residue in the transmembrane region, especially helix 2; the unusual orientation of Arg 159 (homologous to Arg 82 in BR)

Cys 128 to Thr, Ala or Ser or Asp 156 to Ala drastically prolongs the open state (relaxation time of up to 100 s),[74,75] while mutation of Glu 123 to Ala creates faster kinetics. These mutational experiments can be interpreted on the basis of the presence or absence of intrahelical hydrogen bonds. The presence of such an intrahelical hydrogen bond causes a kink in an αhelix, which may slow down the closing of the channel.[64] The conductance of ChR1 is low at physiological conditions, which limits its use in optogenetic experiments.[49,66,67] Compared to ChR2, it has a faster channel kinetics.[51] Volvox ChR1 was initially thought to be a perfect supplement to ChR2 in combination experiments, due to its red-shifted absorption maximum ($\lambda_{max} \approx 530$ nm).[62] However, it is also strongly excitable by 400 nm light, therefore cross-interaction with ChR2 hampers its usability.[71] Furthermore, its expression level is poor.[67]

2.2.3 Photocyle Model for Channelrhodopsin

The different kinetics between ChR1 and ChR2 are supposed to origin in their slightly different role in *C. reinhardtii*: ChR2 generates larger currents and are more easily degraded in high light intensity conditions. Therefore ChR2 is thought to be more involved in dim light conditions, while the faster ChR1 is used predominantly in high light intensity conditions.[51] However, the photocurrents seem to be predominantly caused by ChR1, with only a minor contribution of ChR2.[66] For ChR1, a single cycle model like in BR could not explain all kinetic data, especially the slow recovery back to the dark

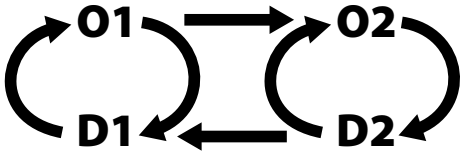


Figure 2.3 Simplified model of a 2 × 2-states photocycle, which could explain the occurrence of multiple absorption maxima as well as the slow recovery kinetics.

state.[76] Instead, with 2 open states and 2 dark states, all observations could be explained, including the occurance of two spectroscopically active species with distinct absorption maxima (463 nm, 505 nm), that co-exist in a pH dependent equilibrium (Fig. 2.3).

Time-resolved spectroscopy on ChR2 characterized 4 intermediates in a single-cycle model, similar to BR (Fig. 1.3).[68,77,78] After photoexcitation of the dark state D470,[c] a red-shifted species, P500, appears within 3 ps, and is thought to be similar to the K intermediate of BR. After that, a blue-shifted intermediate, P390, appears after 10 µs and indicates a deprotonated Schiff base. During the transition from P390 to the next intermediate P520, the channel opens, which happens on a time scale of 2 ms. Channel closing is associated with the decay of the P520 intermediate to the last intermediate P480, which happens within 10 ms. In the last step, P480 relaxes within 20 s to the dark state. Alternatively, there exists photo-cycle models with 2 open states, similar to Figure 2.3.[79,80]

2.3 Optogenetics

First optogenetic approaches were performed in rat neurons and required co-expressions of three separate proteins.[81,82] With the discovery of the channelrhodopsins, optogenetic experiments were now possible via expression of a single protein.[36,83,84] From then, usage of ChRs as optogenetic tools exploded.[85] In 2010, optogenetics was designated *method of the year* by the journal *Nature Methods*,[86–88] and breakthrough of the decade by *Science*.[89]

The general idea of optogenetics is very simple: take a gene of a light-activatable protein that generates some sort of electric or chemical response, combine it with a specific promoter for the cell type, in which the protein is to be expressed, enclose the DNA construct in a virus and infect the tissue with that virus. Using lenti or adeno-associated viruses, mouse, rat and primate brains could be succesfully targeted.[90]

The protein will then be expressed by the cells' ribosomes and inserted into the membrane. The result is, that specific cell types become targetable by light and can be switched on. Action potentials can be generated and further reactions, e.g. changes in behavior of the organism, can be recorded. ChRs depolarize the cell membrane. In combination with halorhodopsin, which hyperpolarizes the membrane, neural activity can be switched on and off by light, which is nicely demonstrated in the transparent worm *Caenorabditis elegans*.[91,92] Besides *C. elegans*, in vivo optogenetic experiments have been performed in a variety of other model organisms, like *Drosophila*, *Zebrafish*, mice and rats. For an overview, the reader is referred to refs. 85,93.

[c] the numbers indicate the absorption maximum in nm of the intermediate. Times reported are taken from ref. 78.

The ultimate goal to which optogeneticists strive towards is the study of neuronal networks in vivo, which may help to investigate psychiatric diseases, such as schizophrenia or autism.[94] First studies for their potential therapeutical usage revealed, that ChRs could restore the transmission of light-induced signals to the visual cortex in a specific form of blindness *(retinitis pigmentosa)*.[95,96] Due to the better spatial resolution, optogenetic implants may be suitable replacements for "brain pacemakers" in deep brain stimulation treatments of Parkinson's disease, chronic pain, dystonia or the Tourette syndrome.[97–99]

In trying to optimize their optogenetic toolbox, neuroscientists engineered a variety of ChR mutants.[71,100] Their success is often based on combinatorial approaches, or *trial-and-error*, rather than conclusions drawn from mechanistic insights. For future advancements, the investigation of the factors, that influence the properties of the ChRs, such as absorption maximum, kinetics and ion selectivity is required. Computer simulations are the perfect tool to achieve that, since structural manipulations can be achieved without great effort and influences on the proteins' properties can be examined on the atomistic level.

2.4 Open Issues

Although, the application of channelrhodopsins in optogenetic experiments reached astonishing levels, unanswered questions can be found in all areas. They can roughly be sorted into two categories. Questions regarding the properties and questions regarding the mechanism. The search for answers to these questions is not only driven by the academic desire to understand, but also by the demand of optogeneticists for engineered ChR variants with desired properties.

An important structural feature with great impact on the properties of the protein is the protonation state of the titrable residues. Spectroscopic experiments of mutants and model pKa calculations indicate the possibility, that in ChR1 and ChR1/ChR2, Glu 162 is protonated, while it is possibly unprotonated in ChR2.[37,64] Computer simulations are able to provide links from the structure to the channel mechanism, so that, with further increase in accuracy and simulation time, the character and function of the counterion can be resolved.

Optogeneticists long for ChR variants, which are red-shifted with respect to ChR2, the main optogenetic tool. This would not only allow for combination experiments, where the cells could be excited with lasers of different wavelengths, but also allows for better tissue penetration.[d] Therefore, an understanding on how the absorption maximum of the retinal is modulated by the protein environment (color tuning) is desirable. Experience gained from the study of other microbial rhodopsins like BR and sensory rhodopsin can be put to good use in the study of the ChRs. In the same spirit, the occurrence of multiple absorption maxima in the spectrum of the ChRs is not understood. It may result from multiple retinal isomers, that co-exist in the ground state. Then, it is possible, that both can be isomerized to open the channel, which would favor a two-state model.

On an even more fundamental note, the question: *Where is the channel?* is still not answered, although computer simulations showed increased water densities and ion binding sites in the transmembrane regions.[37,101] It is of course a major interest of both simulating and experimenting scientists to elucidate the mechanism of the channel activation. Its understanding will guide future "channel

[d] Blue light is also absorbed by other abundant chromophores, such as flavin, hemoglobin or melanin.[100]

engineering" work, since mutation experiments often improve only one property, while sacrificing another. Spectroscopic experiments indicate, that the Schiff base becomes deprotonated during the photocycle. Recent experiments and simulations suggested intriguing differences to BR: The primary proton acceptor in ChR2 seems to be Asp 253 and not Glu 123, which would be homologous to Asp 85, the primary proton acceptor of BR.[78] The reprotonation of the Schiff base occurs via Asp 156, possibly via mediating hydrogen bonds and water molecules. This suggestion explains, why the mutation of Asp 156 drastically slows down the closing of the channel.

Because the average Ph.D. lifetime is too short to address all of the above mentioned issues, the center of attention of this thesis is directed towards the active site of the ChRs and the connection between spectroscopy and structure. In the first computational study on color tuning in ChRs, we analyze the binding pocket and examine the effects of amino acid substitutions on the absorption maximum. With respect to the channel mechanism, we connect the structural model for an important structural motif in the active site, the DC gate, to vibrational and crystallographic experimental data.

Part II

Methods

CHAPTER 3

Quantum-Chemical Methods

Processes in biomolecules, that can be addressed by experiments and/or computer simulations, span different orders of magnitude in both space and time (Fig. 3.1). Ultrafast processes, e.g. the photoisomerization of the retinal in a rhodopsin, usually take place within several hundred femtoseconds. Their description requires an explicit treatment of the electrons and their correlation as accurately as affordable. On the other hand, in order to examine the native state of a biomolecule, the environment of the protein, i.e. solvent and lipid molecules, must be taken into account as well, which easily leads to system sizes of the order of 100 000 atoms. Additionally, to obtain a good equilibrium ensemble of the structure, a simulation has to be performed for as long as possible, which amounts to at most a microsecond on modern computers. This is typically the domain of molecular dynamics (MD) simulations with molecular mechanics force fields. The combination of both worlds gives rise to the idea of quantum mechanics/molecular mechanics (QM/MM) methods.

Several of the methods required to perform simulations on light-activatable proteins, such as Hartree-Fock (HF), density functional theory (DFT) or MD are nowadays considered a standard, on which several excellent textbooks have been written.[102–104] This chapter briefly recapitulates the used quantum-chemical methods for the ground and excited states. Throughout this chapter, the Dirac bracket notation will be used, i.e. $\langle \Psi \left| \hat{H} \right| \Psi \rangle = \int \Psi^* \hat{H} \Psi \mathrm{d}\tau$, with τ being an abbreviation for all the integration variables. Atomic units are used.

3.1 Ground State Quantum Chemistry

As mentioned, ultrafast processes in biomolecules, e.g. electronic excitations, require the investigation of the motion of electrons around the nuclei. In order to do that, one must try to solve the (stationary) Schrödinger equation

$$\hat{H} \left| \Psi \right\rangle = E \left| \Psi \right\rangle \tag{3.1}$$

In typical calculations, only the wave function with the lowest energy, the ground state $\left| \Psi^0 \right\rangle$, or a few higher lying states, the excited states, are considered. The Hamiltonian $\hat{H}$ describes the non-relativistic motion of n electrons around N nuclei. Usually, the Born-Oppenheimer approximation is applied,

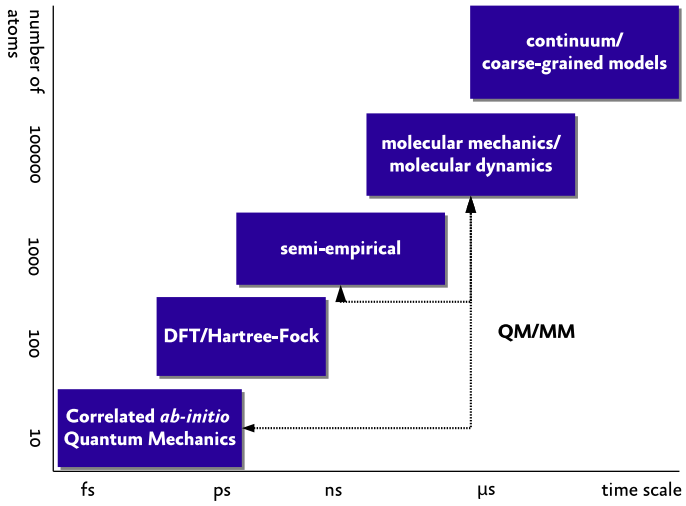


Figure 3.1 Different time and size regimes of processes in biomolecules require fundamentally different approaches of the computational methods. In this work, all of these methods except continuum or coarse-grained models were used, with a focus on QM/MM methods.

which in essence separates the motion of the nuclei from the electrons, so that the Hamiltonian takes the following form:

$$\hat{H}(R,\vec{r}) = \underbrace{-\frac{1}{2}\sum_{i}^{n}\nabla_i^2}_{\hat{T}_e} \underbrace{-\sum_{i}^{n}\sum_{A}^{N}\frac{Z_A}{|R_A-\vec{r}_i|}}_{\hat{V}_{eN}} + \underbrace{\sum_{i,j>i}^{n}\frac{1}{|\vec{r}_i-\vec{r}_j|}}_{\hat{V}_{ee}} + \underbrace{\sum_{A,B>A}^{N}\frac{Z_A Z_B}{|R_A-R_B|}}_{V_{NN}} \tag{3.2}$$

and contains the kinetic energy of the electrons, $\hat{T}_e$, the nucleus-electron interaction, $\hat{V}_{eN}$, the electron-electron interaction, $\hat{V}_{ee}$, and the classical nucleus-nucleus repulsion, V_{NN}, which is a constant. It depends on the electronic degrees of freedom, $\vec{r}$, and *parametrically* on the nuclear coordinates, R. Naturally, an analytic solution is only possible for one-electron systems, such as the H_2^+ molecule. The goal of quantum chemistry now consists in finding the best possible approximation of the wave function, which is guided by the variational principle (eq 3.3). It states that any trial wave function will have an energy expectation value higher than the true ground state energy E_0.

$$E = \frac{\langle \Psi | \hat{H} | \Psi \rangle}{\langle \Psi | \Psi \rangle} \geq E_0 \tag{3.3}$$

3.1.1 The Hartree-Fock method

The simplest idea to construct the many-electron wave function $|\Psi\rangle$ is to take a product of one-electron wave functions, the molecular orbitals (MOs). Unfortunately, this so-called Hartree ansatz fails to fulfill the Pauli antisymmetry principle for fermion wave functions. A substantially better idea is to

take a proper antisymmetrized product of one-electron wave functions, which is immediately satisfied by the Slater determinant.

$$|\Psi\rangle = \frac{1}{\sqrt{n!}} \det |\psi_1 \, \psi_2 \, \psi_3 \cdots | \tag{3.4}$$

In HF theory, the many-electron wave function is approximated by a single Slater determinant. Now, one has to find the best possible set of MO $\{\psi_i\}$, that minimizes the expectation value of the Hamiltonian under the constraint of orthonormality, i.e. $\langle \psi_i | \psi_j \rangle = \delta_{ij}$, which is typically handled by means of *Lagrange multipliers*.[102,105]

Variation of the MO under the orthonormality constraint gives rise to the HF equations:

$$\hat{f}|\psi_i\rangle = \varepsilon_i \, |\psi_i\rangle \tag{3.5}$$

with the definition of the Fock operator $\hat{f}_i$ and the Hartree-Fock potential $\hat{v}^{\mathrm{HF}}$,

$$\hat{f} = \hat{h} + \sum_j^n \underbrace{(\hat{J}_j - \hat{K}_j)}_{\hat{v}^{\mathrm{HF}}} \tag{3.6}$$

which contains the one-electron operator $\hat{h}$, describing the motion of one electron in the "external" potential v^{ext} of all nuclei (charge Z_A at position R_A):

$$\begin{aligned} \left\langle \psi_i \left| \hat{h} \right| \psi_i \right\rangle &\equiv h_i \\ \hat{h} &= -\frac{1}{2}\nabla^2 + v^{\mathrm{ext}} \\ v^{\mathrm{ext}} &= -\sum_A^N \frac{Z_A}{|R_A - \vec{r}|} \end{aligned} \tag{3.7}$$

the Coulomb operator, $\hat{J}_j$, describing the classical Coulomb repulsion of two charge densities:

$$\begin{aligned} \hat{J}_j &= \left\langle \psi_j' \left| \frac{1}{|\vec{r} - \vec{r}'|} \right| \psi_j' \right\rangle \\ \langle \psi_i \left| \hat{J}_j \right| \psi_i \rangle &= \left\langle \psi_i \psi_j' \left| \frac{1}{|\vec{r} - \vec{r}'|} \right| \psi_i \psi_j' \right\rangle \equiv J_{ij} \end{aligned} \tag{3.8}$$

and the exchange operator, $\hat{K}_j$, which arises from the antisymmetry of the Slater determinant and has no classical analogue:

$$\langle \psi_i \left| \hat{K}_j \right| \psi_i \rangle = \left\langle \psi_i \psi_j' \left| \frac{1}{|\vec{r} - \vec{r}'|} \right| \psi_j \psi_i' \right\rangle \equiv K_{ij} \tag{3.9}$$

The Fock operator, via $\hat{J}$ and $\hat{K}$, depends on the MOs $\{\psi\}$, which are also its solutions. Therefore, the HF problem needs to be solved iteratively until the field generated by the electrons is self-consistent (SCF). The Hartree-Fock potential v^{HF} describes the electron-electron interaction in an averaged fashion. In many applications, this averaging needs to be corrected by post-HF methods (cf. 3.2). While

the HF equations (eq 3.5) can be solved numerically, one usually expands the MOs into basis functions (linear combination of atomic orbitals (LCAO)):

$$|\psi_i\rangle = \sum_{\mu}^{\text{basis}} c_\mu^i \, |\mu\rangle \tag{3.10}$$

There are many ways to define a basis set, but two choices are most popular. Atom centered Slater-type orbitals: $|\mu\rangle \propto \exp(-\zeta|\vec{r}|)$ are better motivated physically, while atom centered Gaussian-type orbitals: $|\mu\rangle \propto \exp(-\beta|\vec{r}|^2)$ have computational advantages. In practice, a linear combination of primitive Gaussian functions is used. If the LCAO ansatz is inserted into the Hartree-Fock equations, the Roothaan-Hall equations are obtained:

$$\begin{aligned}
&\mathbf{FC} = \mathbf{SC}\varepsilon \\
&\mathbf{F} = \mathbf{h} + \mathbf{P} \cdot \mathbf{G} \\
&F_{\mu\nu} = \langle\mu|\hat{f}|\nu\rangle = h_{\mu\nu} + \sum_{\lambda\sigma} P_{\lambda\sigma} \; G_{\mu\nu\lambda\sigma} \\
&S_{\mu\nu} = \langle\mu|\nu\rangle \\
&P_{\mu\nu} = \sum_{i}^{\text{occ}} c_\mu^i c_\nu^i \\
&G_{\mu\nu\lambda\sigma} = \sum_{\lambda\sigma}^{\text{basis}} \left(\langle\mu\lambda|\tfrac{1}{|\vec{r}-\vec{r}'|}|\nu\sigma\rangle - \langle\mu\lambda|\tfrac{1}{|\vec{r}-\vec{r}'|}|\sigma\nu\rangle \right)
\end{aligned} \tag{3.11}$$

The Fock matrix **F** contains the matrix of one-electron integrals (eq 3.7), the density matrix **P** and the tensor of two-electron integrals **G**, which may contain up to four different atomic orbital (AO) indices. The matrix **S** is the overlap matrix. **C** contains the set of expansion coefficients, c_μ^i, and the diagonal matrix ε contains the orbital energies. Solving the Roothaan-Hall equations becomes equivalent to diagonalizing the Fock matrix **F**. The most time consuming part is assembling the tensor **G** (effort formally grows with M^4, when M is the number of basis functions). In the vast territory of semi-empirical methods, many elements of **G** are neglected, and the remaining ones are parametrized.

The density matrix is often used for *population analysis* and to introduce the concept of partial atomic charges. The following identity is exploited (n = number of electrons):

$$n = \sum_{i}^{\text{occ}} \langle\psi_i|\psi_i\rangle = \sum_{i}^{\text{occ}} \sum_{\mu\nu}^{\text{basis}} c_\mu^i c_\nu^i \, \langle\mu|\nu\rangle = \sum_{\mu\nu}^{\text{basis}} P_{\mu\nu} \cdot S_{\mu\nu} \tag{3.12}$$

There is no unique way to partition the electron density and assign partial charges. One of the most straightforward ways is the Mulliken population analysis.[106] The Mulliken population of atom A, ρ_A, and the Mulliken partial charge q_A is then given by:

$$\begin{aligned}
\rho_A &= \sum_{\mu\in A}^{\text{basis}} \sum_{\nu}^{\text{basis}} P_{\mu\nu} \cdot S_{\mu\nu} \\
q_A &= Z_A - \rho_A
\end{aligned} \tag{3.13}$$

where Z_A is the nuclear charge of atom A. The straightforwardness of its implementation comes with certain drawbacks, e.g. charge separation is overestimated due to the equal distribution of the off-diagonal elements of $\mathbf{P}$.[102]

3.1.2 Calculation of Vibrational Frequencies

There are at least two wide-spread approaches to compute vibrational frequencies. The *normal mode analysis* requires the computation of the *Hessian* matrix, i.e. the matrix of second derivatives of the energy with respect to coordinates. In cases, where analytical second derivatives are not available, the computation is performed by numerical differentiation of the forces, which quickly becomes cumbersome, since for N atoms, the forces for $6 \cdot N$ coordinate displacements have to be computed. The coordinates in the Hessian are mass-weighted: $\vec{y}_i = \sqrt{m_i}\Delta\vec{x}$, with $\Delta\vec{x}$ being the displacement. The vibrational motions (normal modes) are obtained as eigenfunctions of the Hessian and the vibrational frequencies from the eigenvalues via $\nu_i = {}^1/_{2\pi}\sqrt{\varepsilon_i}$. Normal modes are by construction harmonic and are therefore better suited for fast vibrations, rather than slow collective motions, that often have a large anharmonic contribution. While thermal fluctuations may be approximated by averaging over multiple conformations, normal mode analyses are usually performed at the equilibrium geometry and therefore formally correspond to 0 K. The intensity of the vibrational bands are calculated separately by derivation (or numerical differentiation) of the dipole moments with respect to the normal mode.

A more elaborate approach, which delivers frequency and intensity at the same time and allows for configurational sampling, is based on the Fourier-transform of the autocorrelation function of the dipole moments (FTDAC).[107–109] The intensity of a transition between initial and final states $|i\rangle, |f\rangle$ associated with frequency ω is proportional to the following expression, which is often called *Fermi's golden rule*.

$$I(\omega) \propto \left| \left\langle i \left| \hat{\vec{\mu}}_{if} \right| f \right\rangle \right|^2 \tag{3.14}$$

which is readily obtained by time-dependent perturbation theory, while $\hat{\vec{\mu}}_{if}$ is the transition dipole moment operator. In the *Schrödinger* picture of quantum mechanics, the initial and final states carry the time dependence, while the dipole operator is stationary. For spectroscopy, it is advantageous to switch to the *Heisenberg* picture, where the operator carries the time-dependence, which can be achieved by a unitary transformation. Then, it becomes possible to identify the band intensity with the Fourier-transform of the dipole moment autocorrelation function.

$$I(\omega) \propto \int_{-\infty}^{\infty} \mathrm{e}^{-i\omega t} \left\langle \mu(t) \cdot \mu(t_0) \right\rangle \mathrm{d}t \tag{3.15}$$

The dipole moments are obtained from a time series of classical, quantum-mechanical or QM/MM dynamics. To approximately account for quantum effects of nuclear motions, a simple correction factor is applied to the intensity:[110]

$$Q_{\mathrm{HC}} = \frac{\beta\hbar\omega}{1 - \exp(-\beta\hbar\omega)} \tag{3.16}$$

While this approach allows for the simultaneous evaluation of intensities, frequencies and band shapes and is also easily extendable to Raman spectroscopy (by substituting the dipole moment with the polarizabilities in eq 3.15), it requires sufficient amount of sampling, which restricts the applicable methods to classical or fast semi-empirical quantum-mechanical methods.

3.1.3 Density Functional Theory

The n-electron wave function is dependent on $3n$ (spatial) variables. By virtue of the Hohenberg-Kohn theorem,[111] it was established, that the electron density $\rho(\vec{r}) \equiv \rho$, a function of three variables only, uniquely defines the ground state and its properties, i.e. the energy may be formulated as a *functional* of the electron density. Alas, this theorem does not show how such a functional is to be constructed. In general, the DFT energy can be decomposed into contributions from the external potential v^{ext} (eq 3.7), the classical Coulomb repulsion between two charge densities, J, the kinetic energy of the electrons, T, and the classical nuclear-nuclear repulsion term V_{NN}.

$$\begin{aligned} E[\rho] &= T[\rho] + V[\rho] + J[\rho] + E_{\text{xc}}[\rho] + V_{\text{NN}} \\ V[\rho] &= \int v^{\text{ext}}\, \rho \, \mathrm{d}\vec{r} \\ J[\rho] &= \frac{1}{2} \int v^{\text{H}}[\rho']\, \rho \, \mathrm{d}\vec{r} = \frac{1}{2} \iint \frac{\rho' \rho}{|\vec{r} - \vec{r}'|} \, \mathrm{d}\vec{r}\, \mathrm{d}\vec{r}' \\ T[\rho] &= \sum_i \left\langle \psi_i \left| -\frac{1}{2}\nabla^2 \right| \psi_i \right\rangle \end{aligned} \tag{3.17}$$

While the kinetic energy may be computed from the density alone,[112] performance of orbital-free DFT for molecules is usually poor.[113] Therefore, most applications of DFT apply the Kohn-Sham approximation,[114] which re-introduces molecular orbitals to construct the electron density. The orbitals are a part of an auxiliary system of non-interacting electrons, which gives the correct ground state density ρ_0. The difference between the kinetic energy of the auxiliary system and the *real* interacting system, the exchange and the correlation contribution is packed into an exchange-correlation functional E_{xc}, which has to be approximated.

Now, the energy functional has to be minimized under the orthonormality constraint of the Kohn-Sham orbitals. Then, a set of one-electron equations, the Kohn-Sham equations, are obtained:

$$\begin{aligned} \hat{f}^{\text{KS}} |\psi_i\rangle &= \varepsilon_i |\psi_i\rangle \\ \hat{f}^{\text{KS}} &= -\frac{1}{2}\nabla^2 + \underbrace{v^{\text{ext}} + v^{\text{H}} + v^{\text{xc}}}_{v^{\text{KS}}} \end{aligned} \tag{3.18}$$

The Kohn-Sham potential, v^{KS}, is dependent on the electron density itself, therefore, the Kohn-Sham equations have to be solved iteratively. If only the exact exchange-correlation potential $v^{\text{xc}} = \frac{\delta E_{\text{xc}}[\rho]}{\delta \rho}$ was known, all effects of electron correlation would be contained in a set of one-electron equations. However, it is not conceivable that this is even possible, since correlation is in essence non-local. The

Kohn-Sham orbitals are usually expanded into atomic basis functions (eq 3.10), and a matrix formulation of the eigenvalue problem is obtained. The elements of the Kohn-Sham-Fock matrix are different from the HF case. Especially the Coulomb part is often approximated via the *resolution of the identity*,[115,116] or density fitting approximation, where an auxilliary basis set $\{\omega\}$ is introduced: $\rho \approx \sum_i c_i \omega_i$, which greatly reduces computational cost while not introducing significant errors in most applications, if the auxilliary basis set is sufficiently large.

The total energy may also be formulated alternatively,[113,117] which will be useful in the next section:

$$E[\rho] = \sum_i^n \varepsilon_i - \frac{1}{2} \int \frac{\rho\, \rho'}{|\vec{r} - \vec{r}'|}\, \mathrm{d}\vec{r}'\, \mathrm{d}\vec{r} + E_{\mathrm{xc}}[\rho] - \int v^{xc}\, \rho\, \mathrm{d}\vec{r} + V_{\mathrm{NN}} \tag{3.19}$$

with the help of eq 3.18, the definition of the density: $\rho = \sum_i^n |\psi_i|^2$, the constraint, that the integral of the density should equal the total number of electrons: $\int \rho\ \mathrm{d}\vec{r} = \sum_i \alpha_i = n$ and recalling the following identity:

$$\begin{aligned} \sum_i^n \varepsilon_i &= \sum_i^n \left\langle \psi_i \left| \hat{f}^{\mathrm{KS}} \right| \psi_i \right\rangle \\ &= T[\rho] + \int v^{\mathrm{KS}}\, \rho\, \mathrm{d}\vec{r} \end{aligned} \tag{3.20}$$

In addition to the specification of a basis set (and a suitable auxiliary basis), the user has to choose from an innumerable amount of approximative exchange-correlation functionals. While the basis set is of relatively minor importance, the performance of density functionals varies greatly. In the simplest approaches, the local density approximation (LDA), E_{xc} is only dependent on the electron density, which works well for uniform electron density distributions, such as metals, but not so much in molecules. Functionals of the generalized gradient approximation (GGA) class are dependent on the electron density and its gradient, e.g. the popular PBE[118] or BLYP[119,120] functionals. The most successful exchange-correlation functionals are hybrids between HF and DFT, e.g. B3LYP,[121] the most popular of them all, contains 25 % of HF exchange. Further levels include also the second derivative of the density or contributions from perturbation theory but are not as widely accepted.

3.1.4 Density-Functional Tight-Binding

Tight-binding theories originate from solid state physics. The name refers to the assumption, that the electrons in a crystal are tightly bound to the atoms, from which they originate. Then the electron density can be expressed as a sum of atomic reference densities, ρ^0, and density fluctuations $\delta\rho$:

$$\rho(\vec{r}) \equiv \rho = \rho^0 + \delta\rho = \sum_a \rho_a^0 + \delta\rho \tag{3.21}$$

After insertion of eq 3.21 into the DFT total energy (eq 3.19), separating terms, that only depend on the reference density ρ^0, and expanding the exchange correlation energy $E_{\mathrm{xc}}[\rho^0 + \delta\rho]$ in a Taylor series around ρ^0:

$$\begin{aligned} E_{\mathrm{xc}}[\rho^0 + \delta\rho] = E_{\mathrm{xc}}[\rho^0] &+ \int \left.\frac{\delta E_{\mathrm{xc}}[\rho]}{\delta\rho}\right|_{\rho^0} \delta\rho \, \mathrm{d}\vec{r} \\ &+ \frac{1}{2}\iint \left.\frac{\delta^2 E_{\mathrm{xc}}[\rho]}{\delta\rho\,\delta\rho'}\right|_{\rho^0,\rho^{0\prime}} \delta\rho\,\delta\rho' \, \mathrm{d}\vec{r}\,\mathrm{d}\vec{r}' \\ &+ \frac{1}{6}\iiint \left.\frac{\delta^3 E_{\mathrm{xc}}[\rho]}{\delta\rho\,\delta\rho'\,\delta\rho''}\right|_{\rho^0,\rho^{0\prime},\rho^{0\prime\prime}} \delta\rho\,\delta\rho'\,\delta\rho'' \, \mathrm{d}\vec{r}\,\mathrm{d}\vec{r}'\,\mathrm{d}\vec{r}'' \\ &+ \dots \end{aligned} \tag{3.22}$$

the total energy, after slight rearrangements can be formulated as:[117,122,123]

$$\begin{aligned} E^{\mathrm{DFTB}} &= -\frac{1}{2}\iint \frac{\rho^{0\prime}\rho^0}{|\vec{r}-\vec{r}'|} + E_{\mathrm{xc}}[\rho^0] - \int v^{\mathrm{xc}}[\rho^0] + V_{\mathrm{NN}} \\ &= \sum_i^{\mathrm{occ}} \left\langle \psi_i \middle| \underbrace{-\frac{1}{2}\nabla^2 + v^{\mathrm{ext}} + \int \frac{\rho^{0\prime}}{|\vec{r}-\vec{r}'|}\,\mathrm{d}\vec{r} + E_{\mathrm{xc}}[\rho_0]}_{\hat{H}^0} \middle| \psi_i \right\rangle \\ &+ \frac{1}{2}\iint \frac{1}{|\vec{r}-\vec{r}'|} + \left.\frac{\delta^2 E_{\mathrm{xc}}[\rho]}{\delta\rho\,\delta\rho'}\right|_{\rho^0} \delta\rho\,\delta\rho' \, \mathrm{d}\vec{r}\,\mathrm{d}\vec{r}' \\ &+ \frac{1}{6}\iiint \frac{\delta^3 E_{\mathrm{xc}}[\rho]}{\delta\rho\,\delta\rho'\,\delta\rho''} \delta\rho\,\delta\rho'\,\delta\rho'' \, \mathrm{d}\vec{r}\,\mathrm{d}\vec{r}\,\mathrm{d}\vec{r}'' \\ &= E^0[\rho^0] + E^1[\rho^0,\delta\rho] + E^2[\rho^0,(\delta\rho)^2] + E^3[\rho^0,(\delta\rho)^3] \end{aligned} \tag{3.23}$$

In the original formulation of the density functional tight-binding (DFTB) method, the Taylor series is truncated after the first two terms.[124,125] If second-order terms are included, the DFTB2[a] method is obtained.[122] Including third-order terms results in the DFTB3 method.[126] The first four terms of eq 3.23 are only dependent on the reference density ρ^0 (eq 3.21). In DFTB, this term is approximated as sum of pairwise potentials, which are fitted to reproduce DFT energies, geometries and vibrational frequencies.

$$E^0[\rho^0] = E^{\mathrm{rep}} = \frac{1}{2}\sum_a \sum_b V_{ab}^{\mathrm{rep}} \qquad (a \neq b) \tag{3.24}$$

Therefore, three-center contributions are neglected. The next term, $E^1[\rho^0,\delta\rho]$, contains the DFTB Hamiltonian matrix elements. The Kohn-Sham orbitals (valence electrons only) are expanded in a minimal basis $\{\mu\}$ (eq 3.10):

$$E^1 = \sum_i^{\mathrm{occ}} \left\langle \psi_i \middle| \hat{H}^0 \middle| \psi_i \right\rangle = \sum_i^{\mathrm{occ}} \sum_\mu \sum_\nu c_\mu^i c_\nu^i H_{\mu\nu}^0 \tag{3.25}$$

[a] This name is preferred over the older synonym *self-consistent charge density-functional tight-binding* (SCC-DFTB) and used throughout this thesis

The diagonal elements of the Hamiltonian matrix are chosen to be Kohn-Sham eigenvalues from an atomic DFT calculation with the PBE[118] exchange-correlation functional and off-diagonal elements are subject to a two-center approximation.[125] The basis functions used in these atomic DFT calculations are contracted Slater-type functions:

$$\mu = \sum_{\zeta} \sum_{i=0}^{2} a_{\zeta i} \cdot r^{l+i} \cdot \mathrm{e}^{-\zeta r} \cdot Y_{lm} \tag{3.26}$$

where Y_{lm} are normalized spherical harmonics. The number of primitives with different ζ is 3 for hydrogen, 4 for carbon, nitrogen and oxygen and 5 for the halogens.[127] Compression radii for the orbitals and densities are used, when computing the reference densities, and the Hamiltonian and overlap matrix elements. These elements are pre-computed and stored for each orbital pair ($s-s$, $s-p_\sigma$, $p_\sigma - p_\sigma$, $p_\pi - p_\pi$) and atom on a fine grid of interatomic distances, thus removing the need to calculate any integrals in the DFTB calculation. Just considering E^0 and E^1, the original DFTB method is obtained, which gives good results for the solid state, where the charges are uniformly distributed. However, in molecules, charge re-arrangements need to be considered. This is done by including $E^2[\rho^0, (\delta\rho)^2]$, which is subject to further approximations: the density fluctuation $\delta\rho$ is written as a superposition of atomic contributions $\delta\rho = \sum_a \delta\rho_a$, which are approximated as simple Slater-type exponential functions using the definition of the Mulliken charge Δq_a (eq 3.13), centered at the nucleus at position R_a (monopole approximation):

$$\delta\rho_a = \Delta q_a \frac{\tau_a^3}{8\pi} \mathrm{e}^{-\tau_a |\vec{r} - R_a|} \tag{3.27}$$

The parameter τ_a is related to the width of the fluctuating charge density and to the atomic Hubbard parameter, which in turn is twice the chemical hardness: $\tau_a = {}^{16}/_{5}\ U_a$.[128] The interaction between the fluctuating charge densities is mediated via a function γ_{ab} (eq 3.29),[122,127] which interpolates between two limiting cases. For vanishing interatomic distances, γ_{aa} represents the electron-electron interaction on a single atom, and is described by the Hubbard parameter U_a. For large interatomic distances, the exchange-correlation contribution vanishes and γ_{ab} reduces to the Coulomb interaction between the two fluctuating charge densities Δq_a and Δq_b. Then, the second-order energy can be written as:

$$E^2[\rho^0, (\delta\rho)^2] = \frac{1}{2} \sum_a \sum_b \Delta q_a \Delta q_b \gamma_{ab} \tag{3.28}$$

$$\gamma_{ab} = \frac{1}{|R_a - R_b|} - S(|R_a - R_b|, U_a, U_b) \cdot h \tag{3.29}$$

where S is a rather complicated expression, depending on the Hubbard parameters and interatomic distances.[122,127] Including E^2 gives the DFTB2 method, which is widely used in many applications for inorganic, organic and biomolecules. The second-order approximation is based on one central assumption: the size of the atom, represented by the Hubbard parameter, is the same for all charge states. However, the more charge is accumulated around an atom, the more the effective size increases and vice versa. This effect can be included by making the Hubbard parameter charge dependent, which

is achieved by the third-order term. $E^3[\rho^0, (\delta\rho)^3]$ is approximated in a similar way, while the third-order variation can be rewritten as variation of the γ function.

$$E^3[\rho^0, (\delta\rho)^3] = \frac{1}{3}\sum_a \sum_b (\Delta q_a)^2 \, \Delta q_b \Gamma_{ab} \tag{3.30}$$

The matrix Γ contains the derivative of the γ function with respect to charge, and is calculated in the following way:

$$\begin{aligned} \Gamma_{ab} &= \left.\frac{\partial \gamma_{ab}}{q_a}\right|_{q_a^0} = \left.\frac{\partial \gamma_{ab}}{\partial U_a}\frac{\partial U_a}{\partial q_a}\right|_{q_a} \\ \Gamma_{aa} &= \left.\frac{\partial \gamma_{aa}}{q_a}\right|_{q_a^0} = \left.\frac{1}{2}\frac{\partial \gamma_{aa}}{\partial U_a}\frac{\partial U_a}{\partial q_a}\right|_{q_a} \end{aligned} \tag{3.31}$$

The Hubbard derivative $U_a^d = {}^{\partial U_a}/_{\partial q_a}$ is again pre-computed. Variation of the total energy (eq 3.23) with respect to the MO coefficients, c_μ^i, gives the DFTB Kohn-Sham equations, which is in matrix form:

$$\begin{aligned} &\mathbf{HC} = \mathbf{SC}\,\varepsilon \\ &H_{\mu\nu} = H^0_{\mu\nu} + S_{\mu\nu}\sum_c \Delta q_c \left(\frac{1}{2}(\gamma_{ac}+\gamma_{bc}) + \frac{1}{3}(\Delta q_a \Gamma_{ac} + \Delta q_b \Gamma_{bc}) + \frac{1}{6}\Delta q_c(\Gamma_{ca}+\Gamma_{cb})\right) \end{aligned} \tag{3.32}$$

Since the Hamilton matrix **H** depends on the Mulliken charges, the procedure has to be repeated, until the charges are self-consistent. A further modification of the γ function has been introduced with DFTB3.[126,129] The factor h in eq 3.29 is 1, when atoms a and b are *not* hydrogen. However, if a and/or b is a hydrogen, h becomes:

$$h = \exp\left[-\left(\frac{U_a + U_b}{2}\right)^{\zeta}(|R_a - R_b|)^2\right] \tag{3.33}$$

The parameter ζ is fitted to reproduce the binding energy of a water dimer correctly and greatly improves the description of hydrogen bonds with DFTB. The motivation to modify the γ function for hydrogen is, that in this case, the use of chemical hardness as a measure of the atom size is a poor choice, while it is usually acceptable for other elements.

3.2 Electron Correlation and Excited States

The calculation of excited states of retinal chromophores is computationally challenging, which is addressed by sophisticated combinations of different concepts. Configuration interaction and perturbation theory based methods are fully appreciated in standard textbooks on quantum or computational chemistry.[102,105]

3.2.1 Configuration Interaction

The HF wave function is the best possible single Slater determinant solution for given a basis set. If the wave function is constructed from a linear combination of Slater determinants, the energy can be further lowered. This is the *configuration interaction (CI)* method.

$$\left|\Psi^{\mathrm{CI}}\right\rangle = C_0 \left|\Psi^{\mathrm{HF}}\right\rangle + \sum_{i}^{\mathrm{occ}} \sum_{a}^{\mathrm{vir}} C_i^a \ \left|\Psi_i^a\right\rangle + \frac{1}{4} \sum_{i,j}^{\mathrm{occ}} \sum_{a,b}^{\mathrm{vir}} C_{ij}^{ab} \ \left|\Psi_{ij}^{ab}\right\rangle + \ldots \tag{3.34}$$

where the singly, doubly, etc. excited Slater determinants Ψ_i^a, Ψ_{ij}^{ab} are generated by exchanging virtual and occupied orbitals. If all excitations are considered, the *full CI* limit in the given basis set is reached, which is possible for only a few correlated electrons. In all practical applications, the level of excitation is restricted to single excitations (CIS), double excitations (CID), singles and doubles (CISD) and rarely, higher excitations.

If the variational principle is applied to the CI wave function, in analogy to the HF method (cf. 3.1.1), the expansion coefficients $C_i^a, C_{ij}^{ab}, \ldots$ and corresponding energies are obtained. In addition to the ground state energy and wave function, which is the lowest eigenvalue and eigenfunction ("root") of the CI matrix, higher lying roots, corresponding to excited states, are often desired. The energy differences between the energies then correspond to excitation energies.

In CI, the MO coefficients are not re-optimized and remain on the HF level. If both CI and MO expansion coefficients are optimized, the multi-configurational SCF method is obtained, in which the reference wave function is expanded in a set of many-electron basis functions: $|\Psi\rangle = \sum_I C_I |\Phi_I\rangle$, where $|\Phi_I\rangle$ are either Slater determinants (eq 3.4) or (space- and spin-)symmetry adapted linear combinations of them (configuration state functions). The number of variational parameters rises very rapidly with the number of electrons and orbitals. To save computational effort, the orbital space is partitioned into an inactive space, where the MO are either fully occupied or empty in every Φ_I and an active space, where the occupation numbers vary (Fig. 3.2). This partitioning is not unique and requires a good chemical intuition or experience of the user.

Just as CI adds the description of electron correlation to the HF single reference wave function, multi-reference CI (MRCI), adds correlation to a multi-configurational wave function:

$$\left|\Psi^{\mathrm{MRCI}}\right\rangle = \sum_{I} C_I \left|\Phi_I^0\right\rangle + \sum_{I,i,a} C_{Ii}^a \left|\Phi_{Ii}^a\right\rangle + \sum_{I,i,j,a,b} C_{Iij}^{ab} \left|\Phi_{Iij}^{ab}\right\rangle \tag{3.35}$$

All truncated CI methods are not size consistent, meaning that the energy of a dimer is not twice the energy of the monomer. Considering CISD, up to double excitations are considered in the monomer, but a pair of doubly excited monomers formally corresponds to quadruple excitations (so-called disconnected), which are missing in CISD. Proper size consistency is achieved in the *Coupled Cluster Approximation*, where the wave function is not expanded linearly as in CI, but exponentially.[102] To approximately account for quadruple excitations, the Davidson correction is usually employed.[130]

$$\begin{aligned} E_{\mathrm{CISDTQ}} &\approx E_{\mathrm{CISD}} + \Delta E_{\mathrm{Q}} \\ \Delta E_{\mathrm{Q}} &= (1 - C_0^2)\, E_{\mathrm{CISD}} - E^{\mathrm{HF}} \end{aligned} \tag{3.36}$$

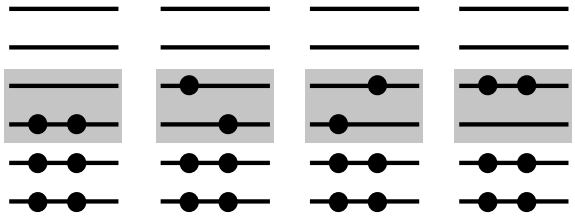


Figure 3.2 Schematic representation of a complete active space with 2 electrons in 2 orbitals, CAS(2,2). The two lowest (highest) orbitals are always fully (un-)occupied. They belong to the inactive orbital space. The two inner orbitals have varying occupation numbers and are considered active. All possible excitations are considered for the active space.

3.2.2 Perturbation Theory

If a system is too complicated to compute, and a closely related reference model system can be found, then the deviation of the real system from the model system can be treated as perturbation.

$$\hat{H} = \hat{H}^0 + \lambda \hat{H}' \tag{3.37}$$

The parameter λ is smoothly varying from 0 to 1. Then, the energies and wave functions also must vary smoothly and can be expanded in Taylor series.

$$\begin{aligned} E &= \lambda^0 E^0 + \lambda^1 E^1 + \lambda^2 E^2 + \dots \\ \Psi &= \lambda^0 \Psi^0 + \lambda^1 \Psi^1 + \lambda^2 \Psi^2 + \dots \end{aligned} \tag{3.38}$$

This can be inserted in the Schrödinger equation and terms of equal order can be collected. In the most popular perturbation theory in quantum chemistry, the MP2 method according to Møller and Plesset, the energy correction is computed up to second order. The reference Hamiltonian $\hat{H}^0$ is chosen to be a sum of Fock operators (eq 3.6): $\hat{H}^0 = \sum_i \hat{f}_i$, the 0th order energies are then the sum of energy eigenvalues $E^0 = \sum_i \varepsilon_i$. The 1st order correction in energy is exactly the HF energy. The 2nd order correction in energy within MP2 can then be calculated as:

$$E^2 = \sum_{i<j}^{\text{occ}} \sum_{a<b}^{\text{vir}} \frac{\left\langle \Phi^0 \left| \hat{H}' \right| \Phi_{ij}^{ab} \right\rangle \left\langle \Phi_{ij}^{ab} \left| \hat{H}' \right| \Phi^0 \right\rangle}{E^0 - E_{ij}^{ab}} \tag{3.39}$$

which can be reduced to an expression containing two-electron integrals. Adding MP2 type corrections to a CASSCF 0th-order wave function leads to the popular CASPT2 method,[131,132] that is often considered to be the "gold standard" for excited states calculations.

3.2.3 Spectroscopy-Oriented Configuration Interaction

As the name SORCI[133] already implies, the method is designed for performing spectroscopic calculations, e.g. to obtain the vertical excitation energy as an approximation to the absorption maximum of a UV/vis spectrum. With that goal in mind, several approximations can be introduced, that would be too crude for computing electronic states, but are appropriate when considering energy differences. The first step consists of the generation of a one-electron basis, usually by performing a HF calculation,

although DFT or multi-configurational SCF calculations could be used as well. For HF orbitals, it is advantageous to re-optimize the virtual orbitals for an $n-1$ electron Fock operator (improved virtual orbitals).[134] Furthermore, orbitals with very high or low energy are discarded from the start.

For the initial definition of a reference space, e.g. a complete active space (CAS) or restricted active space (RAS), it is of utmost importance that all "relevant" orbitals are inside the active space. In the case of polyenes for example, the minimal active space considers all π-electrons and orbitals. After diagonalization of this initial reference space, only those configurations are considered, that have a higher coefficient than a first threshold, T_{pre}. These configurations form the final reference space S and after re-diagonalization yield the 0th order wave function for each state I: $|\Psi_I^0\rangle$.

The next step is an MP2 calculation (eq 3.39) based upon the multi-reference 0th order wave function for all configurations $\{\Phi_\mu\}$, which are not in the reference space S. If this energy is larger than a second threshold, T_{sel}, the respective configuration is assigned to the strongly-interacting subspace R', if lower, it is assigned to the weakly-perturbing subspace R''. In that manner, the size of the CI matrix can be reduced significantly. To further reduce the number of configurations $\{\Phi_\mu\}$ to the *relevant* configurations, i.e. to neglect those, whose contribution cancel upon forming energy differences, the *difference dedicated CI* approach is employed.[135] In that way, the most numerous class of configurations, which contain double excitations from the internal (occupied in all references) to the external (empty in all references) can be omitted.[133]

In a subsequent MRCI step, the Hamiltonian is diagonalized in the space of S and R' to yield energies E_I^a and states $|\Psi_I^a\rangle$ for each state of interest. The multi-reference version of the Davidson correction (eq 3.36) is added to correct the MRCI energies. The energy of each state I is then:

$$E_I = E_I^a\{S, R'\} + E_I^{\mathrm{Davidson}}\{S, R'\} + E_I^2\{R''\} \tag{3.40}$$

To remove any bias, that the initial one-electron basis might have had, state-averaged natural orbitals as eigenfunctions are generated from the reduced density matrix. Thus, a new one-electron basis is constructed, which is substantially smaller than the original one, owing to a third threshold, T_{nat}, which discards all orbitals, whose occupation number is close to 0 or 2. The whole procedure described above is then repeated once again, with a slightly increased DDCI list. In our experience, the bias of excitation energies cannot be eliminated completely by the SORCI procedure, but with "reasonable" choices, the influence is usually minor (cf. 5.1, p 47).

3.2.4 OM2/MRCI

The orthogonalization model 2 (OM2) Hamiltonian[136] is a semi-empirical HF based Hamiltonian, where one-electron three-center integrals $h_{\mu\nu}$ (eq 3.11) and two-electron integrals involving three or four atoms are neglected. The overlap matrix is reduced to a unit matrix and the remaining integrals are parametrized (neglect of diatomic differential overlap, NDDO).[102] In OM2, correction terms are applied, that account for orthogonalization effects. In combination with a specific implementation of MRCI, the graphic unitary group approach (GUGA),[137] excitation energies can be obtained efficiently. While the excitation energy is usually overestimated by ≈0.3 eV compared to SORCI, this overestimation was found to be very systematic for different structures, so that comparisons of relative excitation energies (shifts) are still valid. Due to the efficient implementation, the active space can be chosen larger than in the SORCI method, so that all relevant orbitals are included automatically,

which eliminates the necessity of user interaction. OM2/MRCI is most advantageously used for the simulation of absorption spectrum, where vertical excitation energies are computed from snapshots obtained from a dynamics simulation trajectory.

CHAPTER 4

Force Field Based Methods

The simulation of large systems for longer times is typically the domain of molecular dynamics. On timescales, that routinely exceed hundreds of nanoseconds, and system sizes of several ten thousands of atoms, the interactions between all nuclei and all electrons cannot be described with a quantum-chemical Hamiltonian (eq 3.2). Instead interactions are parametrized in an empirical potential energy function, the so-called *force field*. The dynamics is performed by the numerical integration of the Newtonian equations of motions.

4.1 Force Fields

The idea behind using classical mechanics to describe interatomic interactions is rooted on the concept of locality in chemistry: electrons are *mostly* located near their original nuclei and different functional groups behave similarly, so that *atom types* can be defined.

In this work, the "Chemistry at Harvard Macromolecular Mechanics (CHARMM)" force field was used for proteins[138,139] and lipids,[140–142] which is of the form:

$$E_{\text{CHARMM}} = E_{\text{bonded}} + E_{\text{non-bonded}} + \sum_{\text{residues}} U_{\text{CMAP}}(\phi, \psi) \tag{4.1}$$

$$\begin{aligned} E_{\text{bonded}} &= \sum_{\text{bonds}} K_b(b - b_0)^2 + \sum_{\text{angles}} K_\theta(\theta - \theta_0)^2 + \sum_{\text{Urey-Bradley}} K_{\text{UB}}(S - S_0)^2 \\ &= \sum_{\text{dihedrals}} K_\phi(1 + cos(n\phi - \delta)) + \sum_{\text{impropers}} K_\omega(\omega - \omega_0)^2 \end{aligned} \tag{4.2}$$

$$E_{\text{non-bonded}} = \sum_{\text{non-bonded}} \left\{ \varepsilon_{ij}^{\min} \left[\left(\frac{R_{ij}^{\min}}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{ij}^{\min}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{4\pi\epsilon_0 \epsilon r_{ij}} \right\} \tag{4.3}$$

The potential energy is decomposed into parametrized interactions between two or more different atom types and includes the usual bond (b), angle (θ) and dihedral angle (ϕ) terms. The Urey-Bradley term is used only for some interactions of three bonded atoms A-B-C and depends on the distance S between atoms A and C.[143] The improper dihedral terms are used to describe the interaction between

atoms A, B and D, bonded to a central atom C via the pseudo dihedral angle ω between atoms A, B, C, and D. The CMAP term[139] in eq 4.1 is a pre-computed quantum-mechanical correction to the backbone torsion potential. While the functional form of the force field is common and shared by all the other major force fields used for the simulation of biomolecules, the interaction specific parameters are different and dependent on the procedure used to obtain them.

The non-bonded interactions (eq 4.3), consisting of the Lennard-Jones (12-6) potential for van-der-Waals interactions and the Coulomb interactions between two point charges, are typically computed for all atom pairs within a user-specified interatomic cutoff distance, except atoms that are connected via one or two covalent bonds. The van-der-Waals interactions are typically short-ranged, since they decay according to $\propto R^{-6}$. The Coulomb law decays only with R^{-1} and has a much greater range. The long-range effects are properly accounted for in the particle mesh Ewald (PME) method,[144] where the interaction potential is split into a short-ranged and long-ranged part. The short-ranged part is computed normally, while the long-ranged part is computed in reciprocal space on a three-dimensional grid.

The partial charges q_a of each atom are calculated for individual residues in gas-phase via HF with a small basis set. The method itself would be nowadays considered outdated and it results in dipole moments, that are on average 20 % too large. While this is horrendous for gas-phase calculations, e.g. when compared to a DFT and larger basis sets, this amount of overpolarization is suitable for condensed phase calculations, implicitly building a certain amount of electronic polarization into the force field. The partial charges are residue specific and not protein-specific, i.e. an aspartate on the surface of a water-soluble protein will have the same charge distribution as another one buried deep inside a hydrophobic core of a transmembrane protein, which is a too crude assumption in some applications (cf. 4.5).

4.2 Molecular Dynamics

The motions of all particles $\{i\}$ in an MD simulation are coupled via forces $\vec{F}$, which are determined from the force field (eq 4.1). Therefore, the Newtonian equations of motions $\mathrm{d}\vec{x}_i^2/\mathrm{d}t^2 = \vec{F}_i/m_i$ cannot be solved analytically, and finite difference methods have to be used. The forces that act on a particle at a time t provide accelerations, which are then combined with the positions and velocities to compute new positions. During the timestep Δt, the force is assumed to be constant. In that way, a time series (trajectory) is generated and can be used for further analysis. There are many slightly different implementations of finite difference methods, which all base on the assumption, that the positions, velocities, accelerations, etc. can be approximated as Taylor series. In this work, both the *leapfrog*[145] and the *Velocity Verlet* algorithms are used.[146] Both are reversible and symplectic, i.e. they conserve the total energy of the system. The timestep Δt has to be chosen small enough, so that the trajectory is still smooth, i.e. no particles clash into each other. This is determined by the fastest motion in the system, which are typically bond stretch vibrations involving hydrogen atoms. A standard choice for a timestep, which is also used in this thesis, is 1 fs.

The leapfrog integrator involves evaluations of the velocities at half-timesteps:

$$\begin{aligned} \vec{x}(t+\Delta t) &= \vec{x}(t) + \Delta t \cdot \vec{v}(t+\frac{1}{2}\Delta t) \\ \vec{v}(t+\frac{1}{2}\Delta t) &= \vec{v}(t-\frac{1}{2}\Delta t) + \Delta t \cdot \vec{a}(t) \end{aligned} \tag{4.4}$$

while in the velocity Verlet algorithm, both $\vec{x}$ and $\vec{v}$ are evaluated at the same time:

$$\begin{aligned} \vec{x}(t+\Delta t) &= \vec{x}(t) + \Delta t \cdot \vec{v}(t) + \frac{1}{2}\Delta t^2 \vec{a}(t) \\ \vec{v}(t+\Delta t) &= \vec{v}(t) + \frac{1}{2}\Delta t(\vec{a}(t) + \vec{a}(t+\Delta t) \end{aligned} \tag{4.5}$$

Realistic simulation conditions involve constant temperatures and pressures, which are achieved by thermostats and barostats. Via thermostat algorithms, the simulation system is coupled to a heat bath, allowing heat to flow into and out of the system. To ensure a proper canonical ensemble (NVT, constant number of particles N, constant volume V, constant temperature T), the Nosé-Hoover thermostat is a popular choice.[147,148] Here, the heat bath is introduced as an additional degree of freedom, which may be considered a friction term. The strength of the coupling is determined by the fictional mass of the reservoir, which is closely related to the kinetic energy oscillations between system and reservoir.

Pressure coupling is implemented in a similar way. The volume of the system is treated as an additional degree of freedom and allowed to fluctuate according to the specified external pressure to obtain a correct NPT ensemble (constant pressure instead of constant volume). This is efficiently achieved by the Parrinello-Rahman method.[149,150]

4.3 Hybrid Quantum Mechanics/Molecular Mechanics

The combination of quantum and molecular mechanics QM/MM[151–154] is a sensible approach in two cases. In some problems, the electronic degrees of freedom really are important. Computing UV/vis spectra of a cofactor embedded in a bigger protein, observing reactions that involve rearrangements of electrons, systems, that are too difficult to treat with classical potentials – in all these cases, a quantum-chemical treatment restricted to the important part of the protein is sufficient. On the other hand, in some simulations, that are inherently quantum-chemical, e.g. the flow of charge within a DNA macromolecule, the UV/vis spectrum of retinal in different proteins, etc. the environment surrounding the quantum part really does matter and needs to be included.

There is no unique way of combining both worlds. In this section, the focus is on a specific implementation that is used throughout this thesis, DFTB/CHARMM,[155] which combines the DFTB2 or DFTB3 method (cf. 3.1.4) with the CHARMM force field (eq 4.1). While the bigger part of the system is treated with a force field $\hat{H}_{\mathrm{MM}}$, a small part is treated with a quantum-chemical Hamiltonian $\hat{H}_{\mathrm{QM}}$:

$$\begin{aligned} \hat{H} &= \hat{H}^{\mathrm{QM}} + \hat{H}^{\mathrm{MM}} + \hat{H}^{\mathrm{QM/MM}} \\ \hat{H}^{\mathrm{QM/MM}} &= \hat{H}^{\mathrm{b+vdW}}_{\mathrm{QM/MM}} + \hat{H}^{\mathrm{el}}_{\mathrm{QM/MM}} \end{aligned} \tag{4.6}$$

The difficult part is $\hat{H}^{\mathrm{QM/MM}}$. In many cases, the QM/MM interactions cross a covalent bond, which in proteins is often the C_α–C_β bond of an amino acid. Since the QM region must be a complete molecule, a hydrogen atom is placed between the C_α and C_β atoms to saturate the QM region. The hydrogen atom is only present in the QM part of the calculation. The forces obtained from $\hat{H}_{\mathrm{QM}}$ must then be redistributed over the neighboring atoms.[153]

To avoid an overpolarization of the QM/MM boundary, the charges of the MM boundary atom, C_α are distributed to the neighboring atoms N, H_N and H_α (DIV scheme).[156]

The bonded interactions between QM and MM boundary atoms, as well as the van-der-Waals interactions, $\hat{H}^{\mathrm{b+vdw}}_{\mathrm{QM/MM}}$, are treated on the force field level. The electrostatical interactions can also be treated on the classical level (mechanical embedding), which neglects the polarization of the QM atoms. Usually, it is explicitly included as an additional term in the Hamiltonian (eq 3.32). This approach is called electrostatical embedding.

$$\begin{aligned}\hat{H}^{\mathrm{el}}_{\mathrm{QM/MM}} &= \sum_A^{\mathrm{MM}} \sum_B^{\mathrm{QM}} \frac{Q_A \cdot Z_B}{\left|\vec{R}_A - \vec{R}_B\right|} - \sum_A^{\mathrm{MM}} \sum_i^{\mathrm{electrons}} \frac{Q_A}{\left|\vec{R}_A - \vec{r}_i\right|} \\ &\approx \sum_A^{\mathrm{MM}} \sum_B^{\mathrm{QM}} \frac{Q_A \cdot q_B}{\left|\vec{R}_A - \vec{R}_B\right|}\end{aligned} \tag{4.7}$$

where Q_A is the partial charge of MM atom A, Z_B is the nuclear charge of QM atom B and q_B is the Mulliken charge (eq 3.13) of QM atom B. The approximation on the second line is made in the spirit of DFTB in order to avoid explicit calculation of two- and three-center integrals of the form $\left\langle \mu \middle| Q_A / |\vec{R}_A - \vec{r}_i| \middle| \nu \right\rangle$. In the SORCI/MM scheme, these integrals have to be computed explicitly.

Including the polarization of the QM region by the MM point charges, the QM/MM interactions may be overestimated. Especially at close distances, where the charge densities would overlap, the point charge approximation is too crude. This overestimation can be counteracted by either scaling these interactions or damping the interactions by using a function similar to the γ function of DFTB (eq 3.29).[157]

4.4 Polarization Effects on Excitation Energies

In a standard QM/MM calculation, the QM region is polarized according to the electric field generated by the MM point charges. The response of the MM atoms to the QM charge density is usually neglected. In the context of the simulation of absorption spectra, the neglect of the mutual polarization between QM and MM atoms often leads to overestimations of the excitation energy.[158–160] In rhodopsins, two important effects require consideration. First, during the $S_0 \rightarrow S_1$ excitation, the intramolecular charge transfer within the π-system is large, which makes it especially sensitive to the electrostatic environment. Second, the retinal is surrounded by several aromatic residues, whose charge densities are able to respond on the same timescale as the $S_0 \rightarrow S_1$ excitation.[159,161] In order to describe the mutual polarization, the fixed MM point charges need to be replaced by dipoles, which is achieved in the QM/MMpol framework described below.[161]

The dipoles induced on atom A by the field of permanent charges q and induced dipoles μ of all other atoms is given by:

$$\vec{\mu}_A = \alpha_A \left(\sum_B \underbrace{\mathbf{T}_{AB} \cdot q_B^{\text{perm}}}_{\xi_A^{\text{perm}}} + \sum_C \underbrace{\mathbf{T}'_{AC} \cdot \vec{\mu}_C}_{\xi_A^{\text{ind}}} \right) \tag{4.8}$$

Using this definition, the polarization (induction) energy can be formulated as:

$$E^{\text{pol}} - \frac{1}{2} \sum_A \vec{\mu}_A \cdot \xi_A^{\text{perm}} \tag{4.9}$$

Since the dipole moment on atom A depends on the dipole moments of all other atoms, this equation has to be solved iteratively until the dipoles are self-consistent. In this scheme, the set of permanent charges q_B^{perm} also include the charges of the QM atoms, which are determined by a multistage restraint electrostatic potential fit (RESP).[162]

The partial atomic charges, used in a force field are typically designed for the condensed phase and therefore already account for some polarization effects implicitly, by using an overpolarizing method for their determination, e.g. HF. In order to avoid a double counting of the polarization effects, proper polarization-free charges have to be obtained, which can be achieved by RESP fits to B3LYP/6-311G(2d,2p) gas-phase calculations. The isotropic atomic polarizabilities α_A are obtained from fits against MP2/cc-pVQZ gas-phase calculations.[161] The tensors $\mathbf{T}, \mathbf{T}'$ mediate the interaction of two multipoles: $T = 1/|\vec{r}|, \mathbf{T} = \nabla T, \mathbf{T}' = \nabla \mathbf{T}, \ldots$, where $\vec{r}$ is the distance between atoms A and B.[163] In order to avoid infinite polarzabilities, whenever atoms A and B are within close distance, one of the charge distributions is *smeared:*[164]

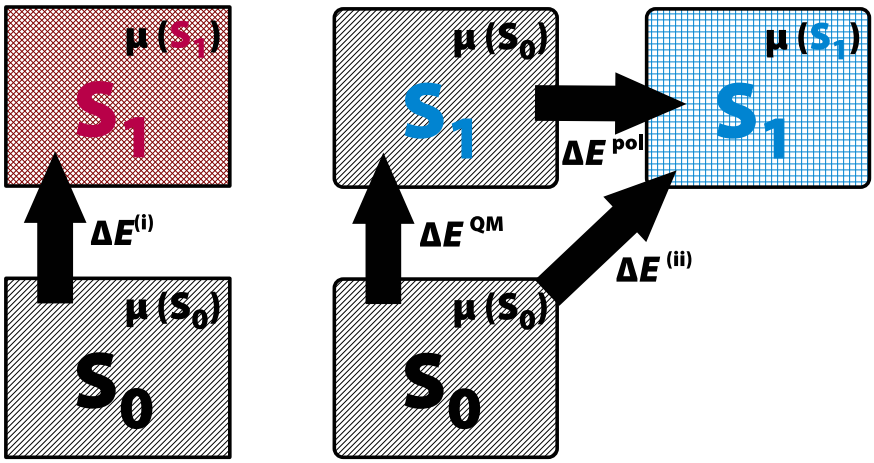
$$\rho = \frac{3a}{4\pi} \mathrm{e}^{-au^3} \tag{4.10}$$

where a is a parameter controlling the damping and $u = |\vec{r}_{AB}|/(\alpha_A \alpha_B)^{1/6}$ is an effective distance between atoms A and B. With that smeared charge distribution, the modified interaction tensor elements in eq 4.8 become: $\mathbf{T} = -(1 - \exp(-au^3))\vec{r}/|\vec{r}|, \mathbf{T}' = \nabla \mathbf{T}, \ldots$.[163]

With SORCI as QM method, charge densities are evaluated for the ground-state S_0 and the (first) excited state S_1, although other states can be computed as well. Then, the dipoles are relaxed separately to give S_0- and S_1-polarized dipoles. Within the QM/MMpol setup, the induced dipole moments are represented as point charges q_1 and q_2, separated by a fixed distance of 0.05 Å from their host atoms,[161] allowing the use of existing QM/MM implementations without further modifications.

$$q_1 + q_2 = q^{\text{perm}} \qquad \vec{\mu} = (q_1 - q_2) \cdot 0.1 \text{ Å} \tag{4.11}$$

Since the charge densities of the S_0 and S_1 states and their transition energy depends on the dipoles of the MM region, and the dipoles relax according to the charge densities, the whole procedure is repeated until the excitation energies are self-consistent. There are two approaches for the calculation of excitation energies under consideration of the mutual polarization in the QM/MMpol framework. In the first approach, the energies of both the S_0 and S_1 states are computed with mutual polarization



(a) approach (i) **(b)** approach (ii)

Figure 4.1 Schematic representation of the two approaches for calculating the $S_0 \rightarrow S_1$ transition energy, including polarization effects in the MM region. **(a)** in approach (i), energies for both states are obtained under mutual polarization. **(b)** in the orthogonality preserving approach (ii), the transition energy is first computed with ground-state polarized dipoles. The dipoles are then relaxed according to the first excited state. The response of the S_1 state to the changed dipoles is neglected (difference between red and blue).

to the respective electron densities. The excitation energy calculated with approach (i) is then just the difference between the two state energies (Fig. 4.1(a)):

$$\Delta E^{(i)}_{S_0 \rightarrow S_1} = E^{\mu(S_1)}_{S_1} - E^{\mu(S_0)}_{S_0} \tag{4.12}$$

Since this approach relies on accurate state energies, the SORCI method cannot be used, since it sacrifices absolute energies to focus on transitions energies. Since $\hat{H}^{\text{el}}_{\text{QM/MM}}$ contains different MM charges $\{Q_A\}$ for the S_1 state than the S_0 state, these states are not orthogonal anymore.[165,166] In order to preserve orthogonality between the S_0 and S_1 states, an alternative approach was designed, where both states are eigenstates of the same Hamiltonian. The procedure of approach (ii) is best described in three steps (Fig. 4.1(b)): (a), obtain the energies of the S_0 and S_1 states for S_0-polarized dipoles $E^{\mu(S_0)}_{S_0}, E^{\mu(S_0)}_{S_1}$. This can be conveniently computed by SORCI as transition energy: $\Delta E^{\mu(S_0)}_{S_0 \rightarrow S_1}$. (b) relax the dipoles according to the S_1 electron density. The energy gain is then $\Delta E^{\text{pol}}_{S_0 \rightarrow S_1}$ (c) the excitation energy for this approach is then given as:

$$\Delta E^{(ii)}_{S_0 \rightarrow S_1} = \Delta E^{\mu(S_0)}_{S_0 \rightarrow S_1} + \Delta E^{\text{pol}}_{S_0 \rightarrow S_1} \left(\mu(S_0) \rightarrow \mu(S_1')\right) \tag{4.13}$$

In this approach, the excitation energy is slightly overestimated, since the response of the S_1 charge density to the changed polarized dipoles is neglected. To remedy this, one can repeat step (b) for S_1 polarized dipoles: $\Delta \tilde{E}^{\mu(S_1)}_{S_0 \rightarrow S_1}$ and relax the dipoles to the S_0 density. Again, the response of the S_0 density to the changed dipoles is neglected, and the excitation energy is underestimated. Averaging both values gives the best estimate for the excitation energy according to approach (ii).

4.5 Limits of Applicability of Standard Force Fields

Polarization effects are not only important for accurately calculating excited state properties,[159–161,167] but show up in a variety of other situations as well. Especially the use of residue-specific, rather than protein-specific charges in classical force field causes erroneous descriptions of structures or energetics. While the focus in the next chapter clearly lies on polarization effects and the lack thereof in simulations with standard[a] force fields, it should not be forgotten to mention that a force field and its parameters is a balanced system, and the atomic partial charges are just one item. The torsional parameters play an equally important role, and much improvement in the performance of force fields was achieved by thorough reparametrizations[168] or numeric correction terms.[139]

In biomolecular simulations, polarization may mean three separate things:[169]

1. *orientational* polarization, which is a re-alignment of a molecule with a dipole moment to an electric field. This effect is usually covered by standard force fields,
2. *geometric* polarization, describing the change in molecular geometry. Changes in molecular geometry due to an electric field are usually rather small in the context of biomolecular simulations, so that this effect can be justifiably neglected in most cases, and
3. *electronic* polarization, which describes the redistribution of electrons. This effect is not properly accounted for in standard force fields.

The electronic polarization of the amino acids is somewhat included in the force field, through the use of overpolarized partial charges. The charges are residue-specific, meaning, a C_β atom will be assigned a different charge in a cysteine or an aspartate, but the C_β atoms of all aspartates will have the same charge, regardless of their specific environment. Obviously, the charge distributions of an aspartate on the surface of a water-soluble protein and an aspartate buried in the hydrophobic core of a transmembrane protein will be different. This effect is neglected in standard force fields, which gives rise to problems in certain situations, despite the overall success of force fields.

4.6 Challenges to Modern Force Fields

Despite all the advances in molecular dynamics simulation methods and the ongoing improvements of standard force fields, there are a couple of problems, where the application of a standard force field may lead to inconsistencies.

A proper accounting for the polarization of the amino acids leads to a stabilization of structures, as shown in many studies. In comparison to a standard force field, a polarized protein-specific force field resulted in an increased stability of secondary structural elements, as indicated by an increase in number and lifetime of hydrogen bonds.[170] Residue-specific order parameters and J-couplings as a measurement for strength and dynamics of hydrogen-bonding interactions, which are calculated with a polarized force field, are also in better agreement to NMR experiments.[171,172]

The inclusion of polarization effects and consequential stabilization of local structure has a direct influence on peptide and protein folding. The sequence dependence of the αhelical content of different

[a] non-polarizable, pairwise additive

substituted polyalanine peptides, as determined by circular dichroism (CD) measurements, could only be reproduced using a polarized force field.[173] A standard force field on the other hand showed no sequence dependence at all. In a notable example, a specific polyalanine mutant was predicted to have a low helical content (≈30 %), which is reproduced by polarized force field calculations, while a standard force field predicts a much higher helical content (≈60 %).[174]

The stabilization of secondary structures by electronic polarization greatly enhances folding dynamics of small proteins as well.[175] Furthermore, a polarized force field greatly stabilizes the native over misfolded, non-native conformations.[176] With standard force fields, the native conformation was found to be significantly higher in free energy than the lowest-energy structure,[177,178] which raises serious questions as to the applicability of such force fields for conformational searches and structure refinement. This is corroborated by a large-scale evaluation of the capability of standard force fields for structural refinement, where it was found, that even for extended simulation times (100 µs), the quality of homology models were not improved at all.[179] The importance of polarization effects for structural stability seems to be independent of the detailed simulation procedure. While most of the above mentioned studies use either implicit or non-polarizable water models, improvements were also achieved, when both protein and first solvation shell are treated with a polarized force field.[180]

An interesting question is whether electronic polarization is important in free energy calculations. As of yet, contradictory findings have been reported in the literature. One study reports, that electronic polarization is a major contribution to the free energy of binding biotin ligands to avidin, which is one of the strongest in nature with $\Delta G = -20$ kcal/mol. The free energy difference between two slightly different ligands is in good agreement with the experiment, if calculated with a polarized force field, while a standard force field predicts almost no free energy difference at all.[181,182] This leads to the previously made conclusion, that electrostatics does not contribute significantly to the binding. Instead, van der Waals interactions being the dominant factor,[183] which is surprising, given the magnitude of the binding free energy. Interestingly, for the same system, a different study comes to the conclusion, that the binding free energies are not much improved over the standard force field.[184] For the binding of different ligands to lysozyme and dodecin, using a polarized force field only marginally improved the binding free energies.[185] It may be concluded, that the polarization effects are different for each protein and whether they are important or not cannot be judged a priori. It is not surprising, that including polarization effects improved scoring of different ligands and different binding poses in docking studies.[186,187] However, it remains questionable, if the inclusion of polarization effects to make predictions more accurate is compatible with the primary goal of docking: high throughput.

Recently, the application of polarized force fields was extended to protein-DNA binding as well.[188] For the binding of a DNA binding domain of a larger protein, calorimetric and spectroscopic experiments predict a binding free energy of -6.0 kcal/mol. A standard force field predicts a positive binding free energy, which means, that this complex is predicted as unstable. Using a polarized force field, the binding free energy becomes negative again, albeit being slightly too negative (-10 kcal/mol).

Polarized force fields also improve the calculations of free energy differences. The difference between protonated and deprotonated species of titrable amino acids buried inside hydrophobic cavities of proteins, the pK_a shift, can be reproduced nicely with a polarizable force field,[189] while standard force fields overestimate this value by a factor of two.[190] However, computing free energies reliably remains a challenging task. The free energy of solvation of small molecules for example depends on the specific choice of a polarized force field.[184,191]

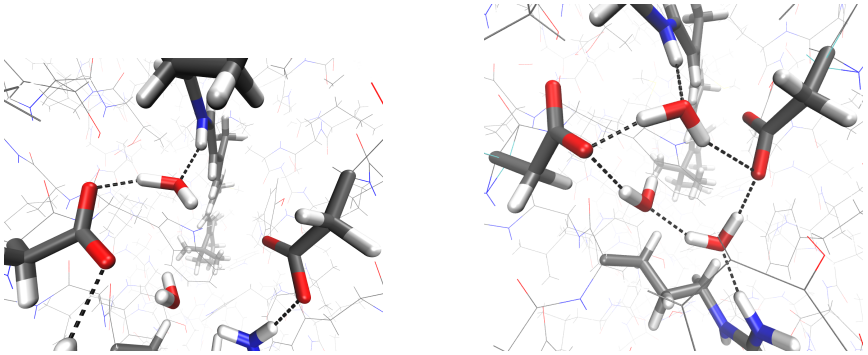


Figure 4.2 Snapshots of the binding pocket of BR from MD simulations. With a standard force field (a), the hydrogen-bonded network gets disrupted, while in QM/MM simulations (b), the pentagonal structure of the water network is nicely retained.

The binding pocket of the proton pumping protein BR features a rigid hydrogen-bonded network, which is not stable in simulations with standard force fields (Fig. 4.2(a)).[192] The experimental structure on the other hand is retained with polarized force fields or in QM/MM simulations (Fig. 4.2(b)). This study especially showed the importance of water molecules, since in their calculations, internal water molecules could have a dipole moment of up to 2.6 D, which is considerably larger than the value of 2.35 D obtained with the classical interaction potential with 3 point charges (TIP3P).[193] Subsequent calculations of vibrational spectra from ensembles obtained with a polarized force field lead to significantly better agreement to experiments,[194] which extends to the flavin proteins as well.[195]

In BR, where the binding pocket structure is well known from many experiments and computer simulations, the extent of the error due to a lack of explicit consideration of polarization effects is apparent from the deviation of the simulated structure from the *real* structure. In case of the channelrhodopsins, the real structure of the binding pocket and detailed information about the interaction of the retinal with its neighboring residues and water molecules is unknown. Due to the even more polar character of the binding pocket in the ChRs, it is easily conceivable, that any conclusions drawn without consideration of this error must be made with outmost precaution.[37] Since the ChRs are ion channels, the location of ion binding sites and their dynamic behavior is of interest in future studies. If ions are just represented as simple point charges, the simulation results will be unreliable.[196–198]

4.7 Strategies to Overcome Limitations of Standard Force Fields

Some authors believe that there is still potential in improving the performance of force fields by better parametrization. They are encouraged by benchmarks, showing that newer parametrizations perform better with respect to selected problems, e.g. NMR J-couplings and chemical shifts.[199] However, for each specific problem, there will always be a special parametrization, that will perform better than the general force field, but it will perform worse in other areas. From a user's point of view, newer parametrizations lack the reliability of the older ones and require careful testing – for the established ones, their limits at least well documented. Furthermore, parametrizations can at best partially compensate a principal lack of physics.

Methods exist, that try to include electronic polarization effects implicitly,[169,200] but it is questionable, whether implicit models will be able to deal with the highly non-uniformity of the charge distribution in heterogenous systems like biomolecules.

Polarizable force fields include dipolar terms explicitly. The induced dipole moment on each atom (eq 4.8) has to be computed in each step of the dynamics, which make them slow. Polarizable force fields have mainly been used in small systems, while they have only been tentatively applied in complex systems like membrane-bound ion channels.[201,202] QM/MM methods include the electronic polarization for a small part of the system only, which in most cases is sufficient, if the QM atoms are selected properly. The main drawback of standard QM/MM implementations is the low computational speed and the consequential restrictions on sampling time. With fragment-based approaches and approximations for the interactions of different fragments, alternative methods open the possibility for faster simulations with QM/MM models.[203,204]

Polarized force fields perform classical simulations, but replace the standard charges of the force field by protein-specific versions.[189,192] If the polarized force field is generated only at the beginning of the simulation, the starting conformation will be stabilized over other, energetically close conformations, which is justified in the context of vibrational spectroscopic simulations, where the required sampling time is short. Situations, which require an unbiased simulation, e.g. the study of the hydrogen-bonded networks in the active site of ChRs, will benefit from a dynamic charge update procedure. While this is an ongoing development, first tests look promising.

Part III

Results

CHAPTER 5

Color Tuning in the Active Site of Channelrhodopsins

Color tuning describes the change of the spectral absorption maximum (λ_{max}) of the retinal with different protein environments. The difference of λ_{max} in organic solvents and different proteins has been termed *opsin shift*.[205]

Modifying the λ_{max} of the ChRs is a desirable goal in protein engineering to extend the possibilities in optogenetic experiments. Being able to excite different cell populations with different light allows for the study of co-operative effects in neural circuits. Further, shifting the λ_{max} to longer wavelengths (bathochromic shift/redshift[a]) allows for deeper tissue penetration and less bleaching of the chromophores.

Color tuning is an excellent application for computational methods, since on one hand it demands a rather high sophistication and therefore is an attractive test for new methods. On the other hand, computer models provide a structural resolution and a detailed breakdown of the different influences on the absorption maximum that is not possible within experimental setups. They not only provide a solid fundament for the interpretation of experiments, but also provide guidance for future mutation experiments.

Parts of this chapter are published as "Color Tuning in Binding Pocket Models of the Chlamydomonas-Type Channelrhodopsins" in the Journal of Physical Chemistry, Part B.[160]

5.1 Calibration of the Computational Approach for the Investigation of Color Tuning

The methods used for the simulation of absorption spectra need to fulfill two different criteria. First, they have to describe the ground-state geometry accurately. Second, the response of the electronic excitation to the electric field, exhibited by the protein environment, need to be included reliably. The first and most important prerequisite is a three-dimensional model of the structure. Since the λ_{max} is very sensitive to variations in the chromophore's geometry, it has to be chosen with care. But from

[a] Here, the terms redshift and blueshift are used as shorter versions of "bathochromic" and "hypsochromic" shift and are entirely unrelated to the red- or blueshifts due to the Doppler effect.

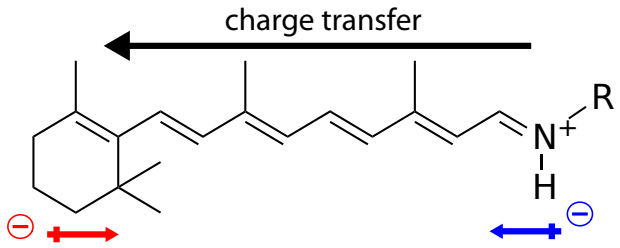


Figure 5.1 Simplified scheme on electrostatic effects of color tuning. During the $S_0 \rightarrow S_1$ excitation, positive charge gets transferred from the Schiff base end to the β-ionone ring. Negatively charged amino acids close to the Schiff base stabilize the positive charge, effectively stabilizing the S_0 over the S_1, causing a blueshift. If they are located close to the β-ionone ring, they stabilize the S_1 state more than the S_0, causing a redshift. For polar amino acids, the orientation – with or against the direction of the excitation-induced charge transfer – is crucial for the direction of the shift.

which of the vast array of methods can be expected, that it reliably reproduces the ground state geometry of the retinal inside the protein binding pocket?

For the first criterion, the ground state geometry, the degree of conjugation in the π-system of the retinal directly influences the λ_{max}. The bond length alternation (BLA), the difference between averaged single and double bond lengths, is directly correlated to the C=C stretch frequency and the absorption maximum.[158,206] DFT methods on the GGA level usually overestimate the conjugation, causing too small excitation energies. HF based methods, including CASSCF, on the other hand yield too large BLA, therefore overestimating the excitation energy of the chromophore. As it is often the case, hybrid density functionals profit from error cancellation and describe the degree of conjugation satisfyingly. As a rule of thumb, admixture of 50 % HF exchange provide the best results (BHLYP or M06-2X functionals).[207,208] The variance in excitation energies calculated from several "good" alternative geometries, like B3LYP (20 % HF exchange) or MP2 is within 0.05 eV.[208] The approximate DFTB provides geometries of similar qualities as GGA DFT, being 0.15 eV red-shifted, but for a negligible computational cost, thus allowing for structural sampling, which is not possible for the more expensive methods.[207] Unfortunately, to straightforwardly determine a "best" method is not possible, since experimental information about the gas-phase structure of the retinal is sparse and debated.[209]

The second criterion, the response of the excitation energy to external electric fields, e.g. the protein environment, is equally important for the study of color tuning. During the electronic excitation from the S_0 to the S_1, the dipole moment of the chromophore changes considerably (≈6 D in gas-phase,[207] 10 D in BR).[158] Therefore, especially charged amino acids can have a large influence on the λ_{max} (Fig. 5.1). The electrostatic effect of the protein can be thought of having two contributions: (a) a large blueshift due to the interaction of the retinal with the negatively charged counterions and (b) an additional shift (either bathochromic or hypsochromic) due to the remaining amino acid, primarily of those within the binding pocket.

Unfortunately, there are not many methods, that are suitable for the computation of the excitation energy of the retinal, especially in a protein environment. Time-dependent DFT methods fail to describe the above mentioned excitation-induced charge transfer on the polyene chain[210] and do not recover the correlation between BLA and excitation energy.[207] Methods, that have been succesfully used to compute the excitation energy are SORCI, CASPT2, NEVPT2, which is closely-related

to CASPT2, but uses a more advanced 0th-order Hamiltonian, quantum Monte Carlo methods and the coupled-cluster based CC2 method. A fair comparison between different calculations is difficult, since seemingly good agreements often result from error cancellations.

For example, it was a popular combination to compute CASPT2 excitation energies from CASSCF geometries, which was rationalized by the very good agreement between the calculated vertical excitation energy (2.00 eV, 620 nm) and experimental λ_{max} (2.03 eV, 610 nm).[211,212] However, it is clear from several studies, that CASSCF geometries differ significantly from other methods.[207,208,210] Especially the BLA is significantly overestimated on the CASSCF level, leading to blue-shifted excitation energies. Further, different 0th-order Hamiltonians for the CASPT2 method yield different results, that leads to shifts of up to 0.3 eV.[208] Therefore, by a fortunate cancellation of errors, the excitation energy itself is still close to the experimental value of 2.03 eV, which validated the use of CASSCF ground-state geometries for the wrong reasons.

The response of the vertical excitation energy to the electric field exerted by the protein and solvent is usually approximated by atom centered point charges, although already the very first QM/MM study accounted for polarization effects approximately.[151] Modern QM/MM schemes, that allow for polarizable embedding of the QM region show promising results.[213] Using point charges, the electronic polarization of the protein in response to the QM charge density and the mutual dispersion is neglected. TD-DFT methods underestimates the shift of the excitation energy due to the presence of point charges by a factor of two, and have therefore been ruled out as useful methods.[207] Multi-reference methods like SORCI or CASPT2 generally agree in the way the point charges affect the excitation energy.

For example, in a QM/MM study, that utilized CASPT2 excitation energies based on CASSCF/AMBER optimized geometries, the excitation energy of BR was calculated to 2.32 eV, while the shift of the excitation energy from gas-phase to protein was 0.3 eV.[214] With SORCI, based on DFTB/CHARMM geometries, we obtain 2.32 eV, while the shift from gas-phase to protein is 0.45 eV (Table 5.4). With respect to the experimental λ_{max} of 2.18 eV, both values are blue-shifted, due to the neglect of polarization of the MM region.[158,160,167,215,216] The polarization effect can be recovered using the *polar.h* protocol (cf. 4.4).

While both excitation energies agree with each other perfectly, the contributions are different. The discrepancy of the protein shifts can be explained by using different sets of point charges, which may lead to differences of about 0.05 eV.[161] Second, we employ a charge-scaling scheme, that mimics the screening effect of bulk water on charged amino acids. This can have an effect of 0.08 eV.[161] Third, it is possible, that different retinal geometries cause different protein shifts with different methods, so that it remains open, if a retinal with a smaller BLA leads to larger responses to external electric fields. However, if one compares the effects of the counterions, the CASPT2//CASSCF/AMBER method predicts 0.93 eV, while our SORCI//DFTB/CHARMM method obtains 0.87 eV, it can be inferred, that SORCI and CASPT2 react similarly to the presence of point charges. On the basis of this comparison, we expect SORCI and CASPT2 to be of comparable accuracy for the calculation of excitation energies in retinal proteins.

Table 5.1 Experimental absorption maxima of the retinal chromophore and selected microbial rhodopsins.

system	λ_{max} in nm	λ_{max} in eV	ref.
retinal in vacuo	610	2.03	212
bacteriodhodopsin	568	2.18	220
channelrhodopsin-1 (pH 8)	463	2.68	66
channelrhodopsin-1 (pH 4.5)	505	2.46	66
channelrhodopsin-2	450/470	2.76/2.64	77
volvox channelrhodopsin-1 (pH 8)	500	2.48	219
volvox channelrhodopsin-1 (pH 6)	540	2.30	219
volvox channelrhodopsin-2	450/470	2.76/2.64	219

5.2 Absorption Maxima of the Channelrhodopsins

The absorption maximum (color) of the retinal can be tuned over a wide range of wavelengths by different protein environments. In light-adapted BR, the λ_{max} is at 568 nm (2.18 eV). The ChRs are blue-shifted with respect to BR (Table 5.1). The absorption spectra of ChR1 and VChR1 show a strong pH dependence, while this is not the case for ChR2. The reason of this is very poorly understood. While a single mutation in ChR1 (Glu87Gln) causes the pH dependence to vanish,[67] other, still unpublished mutations, seem to indicate that Glu 87 is not solely responsible for the protochromism. For ChR1, it was reported, that the absorption maximum at alkaline pH is fine-structured, which vanishes upon acidification.[66] Together with the double absorption maxima in ChR2,[77] this provides strong evidence for two retinal conformations being present in the ChRs. Resonance Raman spectroscopic studies[217] and retinal extraction experiments[218] showed, that in ChR2, the ratio between all-*trans* and 13-*cis* is 70:30. However, action spectra recorded from expressed ChRs, which monitor the wave length dependence of the photocurrent, usually show only one maximum.[50,67] Therefore, the extent, to which different retinal conformations contribute to the function of the retinal, and are therefore relevant for the color tuning is far from unresolved. However, the dominating conformation is all-*trans*, which is the focus of computational studies. VChR1 is the most red-shifted of the chlamydomonas-type channelrhodopsins with a λ_{max} of 540 nm at pH 6. The spectra of VChR2 and ChR2 are virtually identical,[219] so that VChR2 is not considered explicitly in this study.

5.3 Binding Pocket Models of the Channelrhodopsins

In a related study of the color tuning in the microbial rhodopsins BR and sensory rhodopsin 2 (sRII), it was established, that the dominating contribution to the λ_{max} difference between the two proteins comes from differences in the binding pocket. Only 10 amino acids are different in the binding pocket, yet the λ_{max} is shifted by 70 nm. The most important residues are the two counterions, both negatively charged. A different interaction with the counterions accounts for ≈45 % of the shift between BR and sRII. Further 45 % comes from mutations of polar residues in the binding pocket. The rest may be attributed to differences in the chromophore geometry or different interactions with conserved polarizable residues.[158]

To test the importance of the binding pocket for the color tuning in the ChRs and examine, how

the mutations affect the interaction of the chromophore with the counterions, we modeled the color tuning in the binding pockets of the ChRs by starting from the BR X-ray structure and mutating up to 28 amino acids into their respective counterpart of the channelrhododopsins. There are 34 residues with at least one heavy atom within 6 Å of the retinal. These residues comprise the binding pocket.[b] The mutations performed to model the binding pockets were chosen based on published sequence alignments,[50,62] since the X-ray structure was not available at that time. With the availability of three-dimensional structural information for both BR and ChR, both structures can be overlaid to obtain the true sequence alignment. As shown in Fig. 5.2, algorithms have difficulties finding the correct alignment for the first two transmembrane helices, since they differ from other microbial rhodopsins significantly, while helices 3 to 7 can be aligned well. Therefore, the mutations discussed here for helix 1 and 2 have to be put in context of the chosen alignment.

Possible long-range effects on the color tuning are captured by a *perturbation analysis*,[221] where the protein geometry is kept fixed, and the charges of the side chains of each residue are deleted. Thus, the electrostatic influence of this residue can be captured separately from geometric rearrangements. Technically, this is done by "mutating" the residue into a glycine, which also effectively removes the charges of the side chain, but keeps the residue neutral. The electrostatic shift of residue I can then be defined as difference of the excitation energies of the wildtype and the glycine mutant of residue I:

$$\mathrm{ES} = \Delta E^{\mathrm{wt}} - \Delta E_I{}^{\mathrm{gly}} \tag{5.1}$$

which describes the effect of the presence of the side chain of residue I. The difference of electrostatic shifts between a binding pocket model of ChR and BR is then defined as:

$$\Delta\mathrm{ES} = \mathrm{ES}^{\mathrm{ChR}}(I) - \mathrm{ES}^{\mathrm{BR}}(I) \tag{5.2}$$

which accounts for the different electrostatic effects of residue I – either conserved or mutated – between a ChR and BR. In addition to all residues in the binding pocket as defined above (distance < 6 Å), also those amino acids with an $|\mathrm{ES}| \geq 0.01$ eV were considered for the mutations (Fig. 5.4). Two histidines are introduced into the binding pocket region, Asp96His, and Phe208His, which were modeled as neutral, since the corresponding residues in BR, Asp 96 and Phe 208 are neutral as well.

Compared to BR, another titrable residue is introduced into the binding pocket via the Met20Glu mutation. In the alignment underlying the binding pocket models, Met 20 is aligned to Glu 87 in ChR1.[50] In the X-ray structure, it is shown, that Met 20 is instead aligned to Thr 98, so that this particular example is not relevant anymore in the context of color tuning mutations of the ChRs, but its effects are still valid, as will be discusssed below. To compare to experimental spectra, recorded under mildly alkaline conditions, this residue was modeled as charged in the binding pocket models.

[b] There is no unique criterion for defining which residues belong to the binding pocket. However, it seems sensible to restrict the number of amino acids to a "first shell" around the retinal.

```
ChR2      ----------------------CYC-AGWIESRGTNGAQTASNVLQWLAA      60
VChR1     ----------------------CYC-EGWLRSRGTSIEKTIAITLQWVVF      55
ChR1      ----------------------CFC-LAWLKSNGTNAEKLAANILQWITF      99
C1C2                               EKLAANILQWITFALSALCLMF      108
bR        ----------------------TGRPEWIWLALGTALMGLGTLYFLVKGM      32

ChR2      GFSILLLMFYAYQTWKSTCGWEEIYVCAIEMVKVILEFFFEFKNPSMLYL     110
VChR1     ALSVACLGWYAYQAWRATCGWEEVYVALIEMMKSIIEAFHEFDSPATLWL     105

ChR1      ALSALCLMFYGYQTWKSTCGWEEIYVATIEMIKFIIEYFHEFDEPAVIYS     149
C1C2                GWEEIVVATIEMIKFIIEYF
bR        GVSDPDAKKFYAITTLVPAIAFTMYLSMLLGYGLTMVPFGGEQ-------      75

ChR2      ATGHRVQWLRYAEWLLTCPVILIHLSNLTGLSNDYSRRTMGLLVSDIGTI     160
VChR1     SSGNGVVWMRYGEWLLTCPVLLIHLSNLTGLKDDYSKRTMGLLVSDVGCI     155
ChR1      SNGNKTVWLRYAEWLLTCPVILIHLSNLTGLANDYNKRTMGLLVSDIGTI     199
bR        ---NPIYWARYADWLFTTPLLLLDLALLV--DAD-QGTILALVGADGIMI     119

ChR2      VWGATSAMATGY-VKVIFFCLGLCYGANTFFHAAKAYI-EGYHTVPKGRC     208
VChR1     VWGATSAMCTGW-TKILFFLISLSYGMYTYFHAAKVYI-EAFHTVPKGIC     203
ChR1      VWGTTAALSKGY-VRVIFFLMGLCYGIYTFFNAAKVYI-EAYHTVPKGIC     247
bR        GTGLVGALTKVYSYRFVWWAISTAAMLYILYVLFFGFTSKAESMRPEV--     167

                                                          #
ChR2      RQVVTGMAWLFFVSWGMFPILFILGPEGFGVLSVYGSTVGHTIIDLMSKN     258
VChR1     RELVRVMAWTFFVAWGMFPVLFLLGTEGFGHISPYGSAIGHSILDLIAKN     253
ChR1      RDLVRYLAWLYFCSWAMFPVLFLLGPEGFGHINQFNSAIAHAILDLASKN     297
bR        ASTFKVLRNVTVVLWSAYPVVWLIGSEGAGIVPLNIETLLFMVLDVSAKV     217

ChR2      CWGLLGHYLRVLIHE-----------------------------------     273
VChR1     MWGVLGNYLRVKIHE-----------------------------------     268
ChR1      AWSMMGHFLRVKIHE-----------------------------------     312
bR        GFGLILLRSR-AIFG-----------------------------------     231
```

Figure 5.2 Sequence alignment of the channelrhodopsins and bacteriorhodopsin.[50,62] Highlighted in blue are mutations for the binding pocket models, differences between the ChRs are shown in red and underlined residues mark the transmembrane helices of BR. The X-ray structure of the ChR1/ChR2 chimera consists of the first five helices of ChR1 and the last two helices of ChR2.[38] The alignment to BR is different only for helices 1 and 2.

5.4 Simulation Setup

5.4.1 General QM/MM Setup

The coordinates of the heavy atoms of BR were obtained from the X-ray structure, deposited under the PDB code 1c3w.[14] Hydrogen atoms were added using the HBUILD module of the CHARMM program package.[143] Standard protonation states were assumed for titrable residues, except for Asp 96, Asp 115 and Glu 204, which were modeled in their protonated form.[29,222] Harmonic restraints were imposed on all C_α atoms of the backbone and oxygen atoms of the water molecules, that are farther than 12 Å away from the Schiff base. The screening effect of bulk solvent on charged amino acids was included based on a charge scaling scheme.[223] The partial charges of any charged, solvent exposed side chain are scaled down by a factor, that is determined from solutions of the Poisson-Boltzmann equation in two dielectric media ($\epsilon = 1$ and $\epsilon = 80$).

For geometry optimizations and MD simulations the DFTB/CHARMM[155] implementation was used, which has been documented widely elsewhere.[224,225] The QM atoms were treated with DFTB,[122] which yields geometries for the chromophore, that are comparable to hybrid DFT, while being 3 magnitudes faster.[226] The remainder of the protein is described by the CHARMM22[138] force field. In this work, DFTB with (diagonal) third-order terms in charge density fluctuations was used, as described by Yang et al.[129] Two different QM fragments were defined. A small QM region (labeled QM1), which consists of the retinal and the corresponding lysine side chain (Lys 216 in BR) (1 residue, 63 atoms, +1 charge) and a larger QM region (named QM6), which also includes the counterions (Asp 85, Asp 212 in BR) and 3 resolved water molecules (no. 401, 402 and 406 in 1c3w) (6 residues, 89 atoms, -1 charge). The QM/MM boundary was chosen to be between the C_α and C_β atoms of the respective QM residues. The valence of the QM fragment was saturated with a hydrogen link atom, while the electrostatic interactions across the QM/MM boundary were treated according to the *divided frontier* scheme.[156]

To test the stability of the binding pocket models, short QM/MM MD simulations were performed with DFTB/CHARMM with a time step of 1 fs. After heating to 300 K and equilibration, the Nosé-Hoover thermostat[147,148] was applied.

Vertical excitation energies for the transition between the S_0 and S_1 state of the retinal were calculated on two different levels. First, the efficient semiempirical OM2/MRCI[136,137] method as implemented in the MNDO99 program package[227] was applied. It was shown,[158,207] that OM2/MRCI excitation energies are overestimated compared to ab initio methods, but that differences between various proteins are described well. As ab initio method to calculate excitation energies, SORCI was applied[133] as implemented in the ORCA program package.[228] The three thresholds of SORCI method were set in accordance with previous studies.[158,207] The split-valence SV(P) basis set[229] was used, augmented with diffuse s- and p-functions on carboxylate oxygen atoms of anionic residues. As initial one electron basis for the SORCI calculation, averaged approximate natural orbitals were used, resulting from a MRDDCI3 calculation (multi-referential, difference dedicated configuration interaction with 3 degrees of freedom) with a CAS(4,4) reference space after the initial restricted Hartree-Fock calculation and generation of improved virtual orbitals. To account for polarization effects on the excitation energy, the polar.h model was applied (cf. 4.4), which was also shown to be reliable in other retinal proteins.[215,230]

5.4.2 Homology Models of the Channelrhodopsins

The sequences of the ChRs were aligned to the BR according to Fig. 5.2. For the ChRs, only the sequences of the presumptive transmembrane region were used, dismissing the large C-terminal ends. Homology modeling was performed with DeepView[231] and the SWISS-MODEL[232,233] web server with the binding pocket models as template structures. Standard protonation states were assumed on all titrable residues, except Asp 195 (ChR1), Asp 156 (ChR2), Asp 151 (VChR1), which are aligned to Asp 115 in BR and confirmed to be protonated in ChR2.[234] The homology models were optimized with the same settings as described in the general QM/MM setup.

5.5 Single Mutations from Bacteriorhodopsin to the Channelrhodopsins

To examine the effect of each single mutation on the λ_{max}, each amino acid in the binding pocket region is mutated individually. All performed single mutations are listed in Table 5.2. As explained above (Fig. 5.1), negative charges introduced in the vicinity of the Schiff base stabilize the S_0 relative to the S_1 state, causing a blueshift. Likewise a negative charge near the β-ionone ring stabilizes the S_1 state relatively, causing a redshift of the λ_{max}. Therefore, it is not surprising, that the largest shift is induced by the Met20Glu mutation, which can be explained by the negative charge and the position close to the Schiff base (Fig. 5.3). The magnitude of this shift alone is of the same order as the overall shift between BR and the ChRs. Due to its negative charge, it polarizes the surrounding amino acids, which is neglected in the standard QM/MM approach, therefore causing too high excitation energies. If the polar.h model is applied, the charge distribution of the protein is allowed to relax in response to the charge, which consequently reduces the shift of the Met20Glu mutation to 0.20 eV.

As it could already be expected from the color tuning between BR and sRII,[158,221] the effects of single mutations, that do not involve charged residues are rather limited. The residues Ser 141, Thr 142 and Ala 215 have been shown to be significant for the color tuning. In the ChRs, their mutation also leads to shifts of the excitation energy of about the same size.[158] The positions of these amino acids can be seen in Figure 5.3.

Generally, the induced shifts vary only slightly for different QM regions and methods. The semiempirical OM2/MRCI method agrees well with the ab initio SORCI method for small shifts. The differences between the methods are most visible, if the single shifts are summed. For example, in ChR2, the summed shifts are +0.51 eV for OM2/MRCI and −0.17 eV for SORCI (both QM6). The shifts are therefore not additive. The discrepancies between the methods can be traced back to mutations close to the Schiff base region of the retinal (Met20Glu, Asp85Glu, Thr90Cys, Ala215Ser), where the differences in the response of the excitation energy are most pronounced. The effect of increasing the QM region from QM1 to QM6 is threefold. First, a transfer of charge is permitted between the chromophore and its counterions (≈0.1 e). Second, the electron density on the counterions is allowed to be polarized by the chromophore and the surrounding (MM) amino acids. Third, the optimized structure will be slightly adapted, because of the different interaction of the chromophore with the complex counterion, if it is inside the QM region. Combined, these three effects lead to differences between QM1 and QM6 excitation energies and shifts. The charge transfer and polarization of the counterions are described differently on the OM2/MRCI and SORCI levels, resulting in different shifts of excitation energies (QM1→QM6).

Table 5.2 Shifts of excitation energies due to single mutations in the binding pocket region (in eV).

			QM1		QM6	
helix	mutation	protein	OM2/MRCI	SORCI	OM2/MRCI	SORCI
1	Met20Glu	ChR1,VChR1	0.40	0.47	0.38	0.47
	Met20Glu	ChR1*				0.20
	Met20Ala	ChR2	0.00	−0.01	0.01	−0.01
	Phe27Leu	all	0.00	−0.01	0.00	−0.02
2	Val49Ser	ChR1,ChR2	0.04	0.02	0.03	0.01
	Pro50Thr	all	0.02	−0.01	0.01	0.01
	Ile52Gly	all	−0.01	−0.01	0.00	−0.02
	Ala53Trp	all	−0.02	−0.07	−0.04	−0.09
3	Asp85Glu	all	0.06	0.05	0.06	−0.03
	Thr90Cys	all	0.10	0.04	0.10	−0.01
	Leu93Ile	ChR1,ChR2	0.00	0.00	0.00	0.00
	Asp96His	all	0.01	−0.01	0.00	−0.01
4	Met118Thr	ChR1,ChR2	0.00	−0.01	0.00	−0.01
	Met118Cys	VChR1	0.00	−0.01	0.00	−0.01
	Thr121Trp	all	0.03	0.03	0.03	0.02
5	Arg134Lys	ChR2,VChR1	0.01	0.00	0.01	0.00
	Trp138Phe	all	0.01	−0.01	0.01	0.00
	Ser141Gly	ChR1,ChR2	0.12	0.09	0.12	0.08
	Thr142Leu	all	0.03	0.00	0.00	0.03
	Met145Gly	all	0.01	0.01	0.01	0.00
	Ile148Thr	all	0.02	0.00	0.01	0.01
6	Trp178Tyr	ChR1	0.00	0.00	0.00	−0.01
	Trp178Phe	ChR2,VChR1	0.00	−0.02	0.00	−0.01
	Leu181Ser	ChR1,ChR2	0.02	0.00	0.02	−0.02
	Leu181Ala	VChR1	0.01	0.00	0.01	−0.01
	Tyr185Phe	all	−0.03	−0.04	−0.03	−0.05
	Trp189Phe	all	0.02	0.01	0.02	0.02
7	Glu204Ser	all	0.02	0.02	0.02	0.01
	Phe208His	all	0.02	0.01	0.02	0.02
	Val213Leu	all	0.01	−0.01	0.01	0.00
	Ala215Ser	ChR1,ChR2	0.08	0.02	0.08	−0.04
	Val217Asn	all	0.03	0.03	0.00	−0.03
	Phe219Trp	all	0.00	0.00	0.00	−0.02
Σ ChR1 (27 mutations)			0.97	0.70	0.87	0.30
Σ ChR2 (28 mutations)			0.59	0.21	0.51	−0.17
ΣVChR1 (25 mutations)			0.71	0.46	0.63	0.21

* MM region treated with the polar.h model instead of point charges.

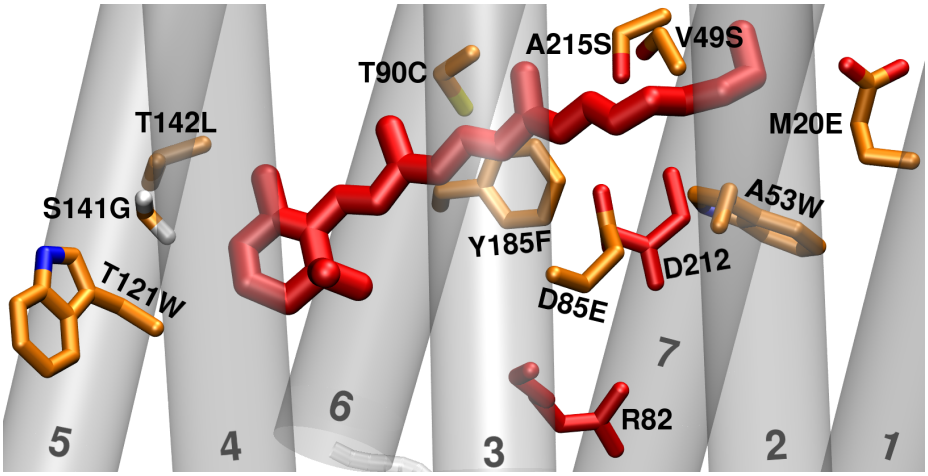


Figure 5.3 Positions of mutated amino acids in the binding pocket of ChR1. Highlighted in orange are mutations with a relatively large impact on the λ_{max}. Highlighted in red are the conserved residues Lys 216, Asp 212 and the chromophore at Lys 216 (numbering from BR).

Table 5.3 Shifts of the vertical excitation energies of the binding pocket models with respect to BR (in eV).

	method	ChR1	ChR2	VChR1
QM1	OM2/MRCI	0.51	0.28	0.40
	SORCI	0.69	0.34	0.42
QM6	OM2/MRCI	0.61	0.38	0.40
	SORCI	0.74	0.46	0.38
expt.		0.50	0.52	0.30

5.6 Perturbation Analysis

To detect amino acids, that are outside the binding pocket, but still have an influence on the excitation energy of the retinal, a perturbation analysis was performed, as described above. Eighteen residues have an ES (eq. 5.1) larger than 0.01 eV (Fig. 5.4), 12 of them being within 6 Å of the Schiff base. Of those 6, that are farther away, Arg 82 and Glu 194 are conserved between BR and the ChRs. Thus, 4 additional mutations are incorporated into the binding pocket models. The mutations are of rather conservative character (Asp96His, Arg134Lys, Thr178Tyr/Phe, Glu204Ser[c]). Since Asp 96 and Glu 204 are involved in the proton-pumping mechanism in BR (p. 7), their mutations indicate functional differences between BR and the ChRs. Indeed, the mutation His173Asp (in ChR1 numbering) leads to a complete loss of light-gated conductance.[49]

5.7 Mutation Induced Shifts of Vertical Excitation Energies of the Binding Pocket Models

The mutations selected due to the close distance to the retinal (24 in ChR1/ChR2, 21 in VChR1) combined with those found in the perturbation analysis (3 in ChR1, 4 in ChR1/VChR1) generate binding

[c] Glu 204 is neutral in BR

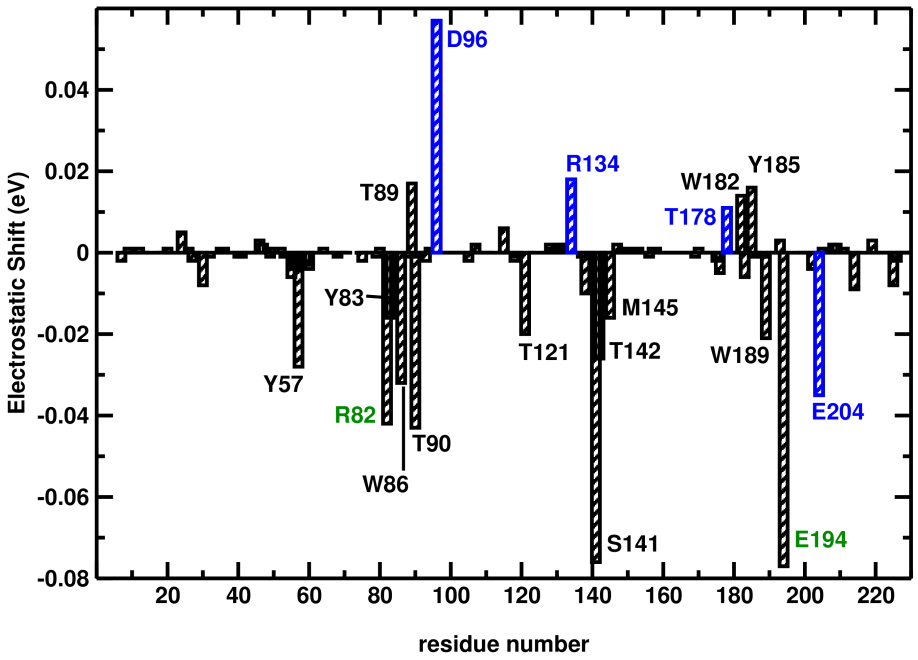


Figure 5.4 Perturbation analysis for BR (QM6 region). Labelled residues have a significant impact on the λ_{max} of the retinal ($|ES| \geq 0.01$ eV). Highlighted in blue are residues, that are outside the binding pocket (as defined by the distance criterion mentioned above) and need to be additionally considered for mutations. Emphasized in green are Arg 82 and Glu 194 which also exhibit a large $|ES|$, but are conserved between BR and the ChRs, and thus omitted from the mutations. The black and labelled residues are inside the binding pocket of BR.

pocket models of the ChRs, which will have shifted excitation energies compared to BR. These shifts are evaluated as differences in the vertical excitation energies of the binding pocket model $\Delta E^{\mathrm{bp}}_{S_1-S_0}$ and bacteriorhodopsin $\Delta E^{\mathrm{BR}}_{S_1-S_0}$.

$$\Delta\Delta E = \Delta E^{\mathrm{bp}}_{S_1-S_0} - \Delta E^{\mathrm{BR}}_{S_1-S_0} \tag{5.3}$$

All binding pocket models are blue-shifted with respect to BR (Table 5.3), which can be expected, considering the type (Fig. 5.2) and position (Fig. 5.3) of mutations. In ChR1, the 27 mutations lead to a blueshift of 0.74 eV (SORCI, QM6). This shift is 0.1 eV larger than the sum of single mutations (Table 5.2), indicating that the effects of mutations are not additive. The main difference between the binding pocket models of ChR1 and ChR2 is the Met20Ala mutation in ChR2 (Met20Glu in ChR1). This is reflected in a smaller blueshift of the binding pocket model of ChR2 (0.46 eV), which is smaller than the effect of the Met20Glu mutation alone, which again signifies the non-additivity of mutation effects. While the Met20Glu mutation is present in VChR1, several other blue-shifting mutations are absent (Ser 141, Ala 215 are conserved, Val49Ala is of conservative character). Thus, the blueshift of the binding pocket model of VChR1 is smallest among the three studied binding pocket models.

For small to medium shifts (ChR2, VChR1), the OM2/MRCI and SORCI methods are in agreement. For ChR1, the SORCI shifts exceed the OM2/MRCI shifts by about 0.15 eV. As shown in Table 5.2, differences between the two methods occur mainly for mutations that are close to the chromophore and perturb the electrostatic environment considerably. The number of such mutations is

Table 5.4 Vertical excitation energies obtained with explicitly polarized MM atoms (SORCI/pol) compared to standard QM/MM (SORCI/MM). Values in eV.

	method	BR	ChR1	ChR2	VChR1
QM1	SORCI/MM	2.32	3.01	2.65	2.73
	SORCI/pol	2.17	2.64	2.50	2.58
QM6	SORCI/MM	2.40	3.14	2.86	2.78
	SORCI/pol	2.18	2.85	2.63	2.61
expt.		2.18	2.68	2.70	2.48

largest in ChR1, and therefore the difference between OM2/MRCI and SORCI is most prominent there.

5.8 Effect of Polarization on the Excitation Energies

Replacing the standard force field charges with the explicitly polarized polar.h model leads to lower excitation energies (Table 5.4). In case of BR, inclusion of the mutual polarization leads to a redshift of 0.22 eV. The perfect agreement with the experimental λ_{max} has to be assessed as a coincidence due to error cancellation. The SORCI method yields vertical excitation energies that are 0.05–0.1 eV below the experimental λ_{max} of gas-phase optimized retinal, if based on DFTB geometries.[207] In addition, dispersion effects are still neglected. Therefore, it can be estimated, that the polarization red-shifts the excitation energy too strongly.

In the ChRs the redshifts due to explicit polarization are similar to BR, which is surprising, considering how much the electrostatic environment of the chromophore is altered in the ChRs. This finding underlines the difficulty of estimating the effect of polarization a priori, which was also demonstrated in the case of the O-state of BR,[230] the last intermediate of the BR photocycle: the excitation energy of the O-state is hardly lowered by explicit polarization, while the lowering is sizable for the BR ground state (Table 5.4), although both structures are similar. Thus for the calculation of abolute values of excitation energies, the inclusion of explicit polarization is vital, supporting the results of previous studies,[159,161,230] but shifts of excitation energies are hardly affected.

5.9 Contribution of Different Retinal Geometries to the Overall Shift

The overall shifts (Table 5.3) result from several distinct contributions. The first part to be analyzed is the different geometry of the retinal in BR on one hand and in the ChRs on the other. By calculating the excitation energies of the retinal in gas-phase (QM1 without protein point charges), one can directly infer the shifts that result from different retinal geometries (Table 5.5). To rule out structural differences between QM1 and QM6 optimized structures, these excitation energies were calculated with the QM6 optimized structure. As stated above, our reference value for the unperturbed chromophore obtained with SORCI for the DFTB optimized 6s-*trans*-all-*trans*-retinal is 1.93 eV. This value agrees well with the experimental absorption maximum of the retinal at 2.03 eV.[212]

Table 5.5 Excitation energies of the chromophore (QM1) computed without external point charges with the SORCI method and structural parameters of the retinal.

	BR	ChR1	ChR2	VChR1
ΔE (in eV)	1.87	1.96	1.95	1.96
planarity* (in °)	10.06	4.96	6.16	8.51
BLA (in pm)	5.39	7.63	6.93	7.18
H-bond to:†	W402	E85	E85	W402

∗ averaged deviation of the dihedral angles from 180°

† hydrogen bonding partner of the Schiff base.

In the ChRs, the different retinal geometries lead to shifts of about 0.1 eV with respect to BR. This shift is slightly larger than between BR and sRII (0.05 eV).[158] In the ChR1/ChR2 chimera, the shift due to a perturbed retinal geometry, obtained with RI-CC2[235,236] has been determined to be 0.05 eV as well.[213]

Among the ChRs, the retinal geometry is similar, so that they also have similar excitation energies for the bare chromophore. One important parameter that correlates with the λ_{max} is the bond length alternation. In the binding pocket models, the BLA increases, as does the overall shift of the excitation energy. The shifts and structural parameters of the retinal (Table 5.5) indicate very similar geometries, implying that the different retinal geometries are just a small contribution to the large overall shift. This in turn implies that the BLA is not the reason for the large blueshift, but a consequence of the electrostatic interaction with the amino acids of the binding pocket. An increased BLA indicates a more pronounced double or single bond character. This is also reflected in the increased planarity of the dihedral angles, especially around the formal C=C double bonds of the retinal backbone. Visible structural differences are restricted to the β-ionone ring and a different conformation of the lysine side chain, which is connected to a different orientation of the N-H bond of the Schiff base.

5.10 Contribution of the Counterions to the Overall Shift

The complex counterion in BR is composed of Asp 85, Asp 212 and the three water molecules (401, 402, and 406) between these residues. One of the counterions, Asp 85, is replaced in all ChRs by a glutamate. The single mutation Asp85Glu induces only minor shifts of about 0.03 to 0.06 eV (Table 5.2) which is of the same order as the shift due to the different retinal geometry. This is supported by the reverse mutation Glu162Asp, which shifts the excitation energy of the ChR1/ChR2 chimera by −0.07 eV. Although being a different type of spectrum than a UV/Vis spectrum of purified protein, the action spectrum of ChR2, expressed in oocytes of *Xenopus laevis*, shows a similar shift for the Glu123Asp mutation. In BR, the two counterions impact the excitation energy to similar extent. The effects are not additive, since the replacement of both residues lead to a different shift than the sum of the single mutations.

Overall, the electrostatic shifts of the complex counterion increases in all ChRs (Table 5.6). Compared to BR, the ΔES (eq 5.2) ranges from 0.06 eV in VChR1 to 0.25 eV in ChR1, with ChR2 being in the middle (0.11 eV). This correlates with the overall trend: the binding pocket models of VChR1 and ChR2 having similar shifts, and ChR1 being the most blue-shifted. By comparing the magnitude of the

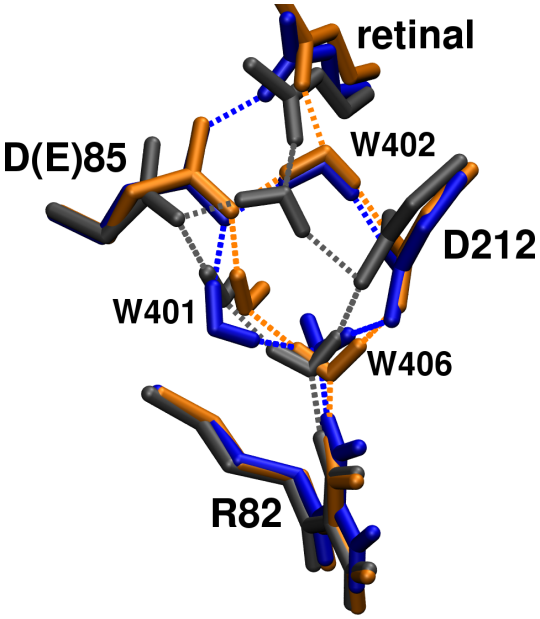


Figure 5.5 Different configurations of the hydrogen bonded network of the retinal with its counterions in the binding pockets; BR = grey, ChR1/ChR2 = blue, VChR1 = orange

Table 5.6 Excitation energies of the chromophore in vacuo, embedded in the protein and with deleted charges of the counterions (in eV). Calculated with SORCI for the QM1 region.

	BR	ChR1	ChR2	VChR1
chromophore in vacuo	1.87	1.96	1.95	1.96
chromophore +HBN*	2.80	2.95	2.92	2.79
chromphore, embedded in the protein	2.32	3.01	2.65	2.73
charges deleted of:				
−Asp (Glu)85	1.92	2.40	2.15	2.33
−Asp 212	1.97	2.71	2.51	2.60
−CCI†	1.66	2.12	1.91	2.03
ΔE(protein) − ΔE(−CCI)	0.66	0.89	0.75	0.70

∗ hydrogen-bonded network, QM1 in the presence of point charges of Asp/Glu 85, Asp 212, Arg 82, waters 401,402, and 406 only

† complex counterion, Asp 85, Asp 212, waters 401, 402, and 406.

Table 5.7 Summed differences of the electrostatic shifts (ES) for the binding pocket models (QM6, OM2/MRCI, in eV). Note that the summation excludes the counterions (Glu 85 and Asp 212), since they are discussed separately.

	ChR1	ChR2	VChR1
$\Sigma\,\Delta$ES (all residues)	0.45	0.10	0.26
$\Sigma\,\Delta$ES (binding pocket)*	0.44	0.13	0.23
$\Sigma\,\Delta$ES (b.p. mutations)	0.48	0.23	0.27
$\Sigma\,\Delta$ES (b.p. conserved)	−0.03	−0.10	−0.04

∗ 38 residues out of 227; 32 residues within 6 Å of the retinal (excluding both counterions) and 6 residues with large ES in the perturbation analysis in BR.

ΔES of the counterions (Table 5.6) with the overall blue-shift (Table 5.3), it is shown that the increased interaction of the Schiff base with the counterions is a major contribution to the overall shift.

If the charges of the counterions are removed (−CCI), the effect of the remaining amino acids are still included in the excitation energy value. In BR, the removal of the complex counterion leads to a redshift, which was also reported previously.[158,214] The excitation energies of the ChR binding pocket models without the counterions' charges are close to the gas-phase value of the retinal, suggesting, that in these models, the binding pocket residues do not cause a redshift. The hydrogen-bonded network causes a blueshift of about 0.9 eV in BR,[214] which is also valid for ChR1 and ChR2, while this shift is smaller in VChR1.

The different electrostatic shifts of the two counterions in the ChRs are correlated to structural differences. On one hand, an increased distance between the retinal and the Asp 212 leads to lower ES of this residue in the ChRs. On the other hand, a shorter distance between the retinal and Glu 85 is accompanied by a higher ES. In VChR1, the distance between the retinal and the Glu 85 is similar as in BR, suggesting a similar ES.

These different distances also correlate with different orientations of the N-H bond of the Schiff base (Fig. 5.5). In BR, and VChR1, the N-H bond is oriented to a water molecule (W402 in BR X-ray structure 1c3w) resulting in similar electrostatic shifts of the counterions. In ChR1 and ChR2 on the other hand, the N-H bond is oriented towards one of the O_ϵ-atoms of the Asp 85 residue, leading to higher electrostatic shifts.

These data are assembled from QM/MM optimized structures. During short QM/MM MD simulations the increased flexibility of the interaction of the Schiff base with its complex counterion becomes apparent. The hydrogen bonding pattern of the BR starting structure seems to be conserved, albeit with increased flexibility of W401. While in BR the hydrogen bond between the Schiff base and W402 is highly conserved during the QM/MM MD, it switches from the Asp 85 counterion to the water W402 back and forth. This switching of the interaction mode of the Schiff base in a more flexible binding pocket could be a hint towards the appearance of multiple absorption maxima in the spectra of the ChRs, as shown for ChR2 with two λ_{max} separated by 0.12 eV (20 nm).[77]

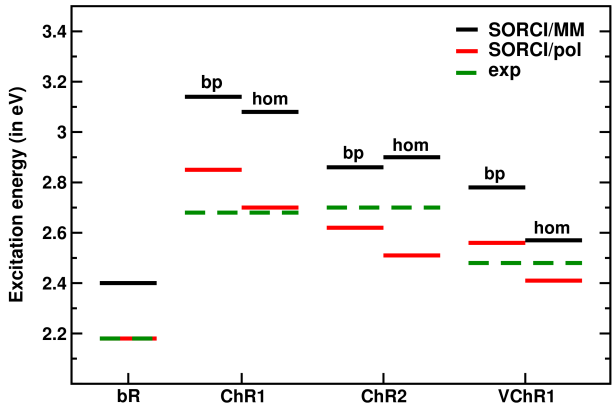


Figure 5.6 Comparison of vertical excitation energies of the homology models with binding pocket models.

5.11 Contribution of the Binding Pocket Amino Acids

To analyze the contribution from the protein environment, a perturbation analysis is performed, as discussed above. The results of this analysis are shown in Table 5.7. The electrostatic shift differences between ChR1 and BR for all amino acids (excluding the counterions) sum up to 0.45 eV ($\Sigma\Delta$ES all residues). This is almost identical to the summed shift of the binding pocket residues ($\Sigma\Delta$ES binding pocket), indicating that the amino acids outside the binding pockets have the same electrostatic shifts as in BR, thus not contributing to the overall shift.

The major part of this shift is caused by the 27 mutations in the binding pocket (0.48 eV), while differences of the conserved residues (Glu 194, Asp 115, Thr 89, Arg 82 and Tyr 57 with largest ΔES) cause a small redshift of (-0.03 eV). This is due to a slightly different orientation of these residues in the binding pocket models compared to BR. In ChR2, where the Met20Glu mutation is replaced by Met20Ala, the binding pocket mutations lead to a summed ΔES of 0.23 eV. Interestingly, the effect of conserved residues is larger than in the other ChRs, causing a redshift of -0.1 eV. The majority of this shift can be traced back to Tyr 57, Asp 115 and Glu 194. Although the coordinates of these residues are nearly identical in the binding pocket models of ChR1 and ChR2, their effect on the excitation energy of the chromophore is different.VChR1 again is similar to ChR1. As mentioned before, several mutations are missing (Leu 93, Ser 141, Ala 215 are conserved) or conservative (Val49Ala, Leu181Ala), so that the blueshift due to the mutations in the binding pocket is only 0.27 eV. The effect of the conserved binding pocket residues is similar to ChR1.

5.12 Estimate of the Contribution of the Rest of the Proteins – Evaluation of Homology Models

By mutating a limited amount of amino acids surrounding the retinal, the binding pockets of the ChRs can be modeled and their effect on the λ_{max} of the chromophore can be estimated as described above. By using these binding pocket models as templates for homology modeling, the (electrostatic) effect of the rest of the protein can be estimated. These estimations have to be interpreted carefully however, since the overall structure of the homology models is still very BR-like. In this manner, the effect of

changing the protein around the binding pockets can be extracted.

It is shown in Figure 5.6, that the excitation energies of the homology models of ChR1 and ChR2 are similar to those obtained for the binding pocket models (Table 5.4), thus leading to blueshifts of similar magnitude. For VChR1, the excitation energy (QM6) is 0.2 eV lower than for the binding pocket model. The reason for this deviation is a larger structural rearrangement during the QM/MM optimization of the VChR1 homology model. For the homology models of ChR1 and ChR2, these rearrangements are smaller, resulting in similar excitation energies as for the binding pocket models.

According to the primary sequences of the ChRs several charged residues are introduced into the N-terminal regions. The exact location of these charged residues highly depends on the sequence alignment, which makes an exact determination of their contribution to the color tuning in the ChRs difficult. In this particular alignment (Fig. 5.2), several of the additional charged residues appear in the loop between helices 2 and 3, rendering them solvent exposed and thus screening their charges. Those amino acids, that are in the helical region face into the membrane region or appear paired with oppositely charged residues thus nullifying their effect on the retinal. For this alignment, only 2 glutamates (ChR1: Glu 122, Glu 123; ChR2: Glu 41; VChR1: Glu 77, Glu 78), but up to 4 positively charged residues (Lys 80, Lys 88, Lys 115, Lys 132; ChR2: Lys 76, Lys 93; VChR1: Lys 44, Arg 71, Lys 88) are introduced in helix 2. The summed electrostatic shifts of these residues is smaller than 0.01 eV, meaning that for this alignment, the additional charged residues do not contribute significantly to the color tuning. However, the appearance of more positively than negatively charged residues is questionable for a cation-conducting channel.

5.13 Updates Since the Publication of the X-ray Structure

We performed the first computational study on the channelrhodopsins, which was published in late 2011,[160] without knowledge of the three-dimensional structure of the protein. We modeled the color tuning in the binding pockets of the ChRs based on the structure of BR. In the literature, multiple sequence alignments for the ChRs exist,[50,51,56] each with its own advantages and disadvantages, differing mainly in the putative transmembrane helices 1 and 2. Algorithms had difficulties predicting these, due to the many charged residues in the ChRs, which is different from other microbial rhodopsins. In fact, most programs for the prediction of secondary protein structures from sequences predict ChR to have only four or five transmembrane helices.[37] Interestingly, a visual inspection of the sequence lead to the discovery of a seven-residue periodicity of charged/polar groups, which made the generation of a structural model possible, where the charged/polar residues are aligned on top of each other in a quasi one-dimensional chain.[37] This computational model of ChR2 is remarkably similar to the X-ray structure, published at the same time.[38,64]

The unique sequence alignment, as determined by the overlay of the X-ray structures of BR and the chimeric ChR1/ChR2 construct, shows, that helices 1 and 2 of the alignment we chose for the study of color tuning[50] are shifted by 12 positions (Fig. 5.2). Therefore, the mutations on these helices, that were selected based on this alignment, are not relevant for the color tuning of the ChR. The mutations, that would have to be considered instead are listed in Table 5.8.

While the structure of the binding pocket model would naturally be different, especially in the Schiff base region, much of the analysis performed in our study is still valid. Mutations of the amino

Table 5.8 Replacements for the mutations on helices 1 and 2 of ChR1 based on the X-ray structure for the ChR1/ChR1 hybrid.

helix	amino acid in BR	*old* alignment[50]			*true* alignment[38]		
		ChR1	ChR2	VChr1	ChR1	ChR2	VChR1
1	Met 20	Glu	Ala	Glu	Thr	Ala	Val
	Phe 27	Leu	Leu	Leu	Cys	Leu	Cys
2	Val 49	Ser	Ser	Ala	Val	Val	Val
	Pro 50	Thr	Thr	Thr	Ala	Cys	Ala
	Ile 52	Gly	Gly	Gly	Ile	Ile	Ile
	Ala 53	Trp	Trp	Trp	Glu	Glu	Glu

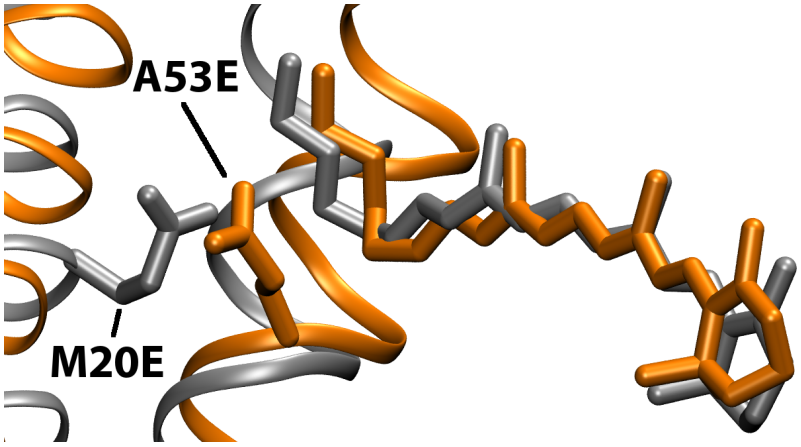


Figure 5.7 Comparison between the Met20Glu mutation performed in this study (grey) and the putative Ala53Glu mutation, that would have to be performed for the channelrhodopsins (orange, Glu 129 in ChR1, Glu 90 in ChR2).

acids Phe 27, Val 49, Pro 50, Ile 52 all had only a small effect on the λ_{max}. With the updated alignment, they are either conserved or conservative (Pro 50 Ala/Cys, Phe 27 Cys/Leu), so that no significant changes are expected with respect to the previous analysis. Interestingly, instead of the Met20Glu mutation, the Ala53Glu mutation introduces a glutamate into the binding pocket in a very similar position (Fig. 5.7), so that the effects of the Ala53Glu mutation should be similar to the Met20Glu mutation discussed above. Given the *old* alignment (Fig. 5.2), the Met20Glu mutation only was present in ChR1 and VChR1. The absence of an additional negative charge in the binding pocket model of ChR2 was responsible for their relatively low blueshift compared to ChR1 (0.52 eV for ChR2 vs. 0.74 eV for ChR1, Table 5.3). This finding corroborates our conclusion, that especially an additional charge introduced in the binding pocket region is able to account for the large blueshift of the ChRs compared to BR, and that other amino acids have a limited impact.

Since the publication of our binding pocket analysis,[160] two studies, that uses the x-ray structure of the chimeric ChR1/ChR2 have been published. In a work, that focusses on the electrostatic effects of each amino acids,[237] similar to the perturbation analysis described above (cf. 5.6), the findings of our binding pocket study are further corroborated, additionally validating the approach to use bind-

ing pocket models for the study of color tuning. Overall, the effects of amino acid substitutions are described in a very similar fashion, if one uses either the full protein[237] or the binding pocket as basis.[160] The mutation Ser295Ala for example, which is the reversion of Ala215Ser of our binding pocket analysis leads to relatively large redshifts and at least partially accountable for the spectral difference between ChR1 and VChR1. This position, directly preceding the retinal binding site (sequence-wise) was also found important in the color tuning of other microbial rhodopsins.[158] In the homology models of the ChRs the charged amino acids outside the binding pocket did not cause additional shifts of the λ_{max} (Table 5.7, p 61). In the X-ray structure, the largest shift could be observed by the mutation Lys209Asp, which is about 15 Å away from the retinal. But even this most drastic change in character leads to an electrostatic shift of only 0.1 eV, thus supporting our conclusion, that the binding pocket is the dominant factor.

In a second study, that uses a different methodological approach, the results support our findings as well.[213] Based on the approximate coupled-cluster method RI-CC2,[235,236] the QM fragment is embedded into the MM region in a polarizable fashion, that allows the mutual polarization of QM and MM regions also for the computation of excitation energies (PERI-CC2).[238] Using this approach, they observe polarization shifts of about -0.21 eV for the chimeric channelrhodopsin, which is similar to the binding pocket model of ChR1 (-0.29 eV) and BR (-0.22 eV), which we computed with the SORCI/polar.h method. [239]

5.14 Conclusions

To further expand their optogenetic toolbox, neuroscientists have created a demand for modified ChR variants. The most successful modifications change either the kinetics of the photocycle or cation selectivity.[69,74,75,239] They show, that useful effects can be achieved by single amino acid mutations.[239] Red-shifted ChRs are desirable for mainly two reasons: for in vivo applications, a ChR with an absorption maximum shifted to higher wavelengths (> 520 nm) would significantly reduce scattering losses and bleaching effects.[240] Moreover, ChRs absorbing light of different wavelengths allow for combinatory experiments, that selectively activate different cell types with lasers of different wavelengths. Is this goal achievable with single mutations or a limited number of them? All computational studies agree, that the effect of single mutations on the λ_{max} is limited and provide a rather fine tuning mechanism. So far, engineering color-shifted ChRs by mutations has not been achieved. The main reason for this is simple: the most promising targets for color tuning are naturally charged residues. Especially in helix 2 there are a number of negatively charged glutamates, whose mutation into a neutral residue could indeed lead to a redshift. Unfortunately, it is precisely these charged amino acids, that are different from other microbial rhodopsins and are therefore crucial for the unique channeling mechanism. Changing these amino acids would alter the channel function. Another factor, which is not understood, is how mutations affect the expression level of the protein, which is at least as important as the intrinsic spectral and kinetic properties of the channel.

The most promising approach so far is the generation of chimeric proteins, where helices of two channelrhodopsins are combined. While VChR1 is a potentially interesting optogenetic tool with the desired long-wavelength λ_{max} of 540 nm (2.3 eV), its expression level and photocurrents is so weak, that it is useless for optogenetic experiments. Swapping the first two helices of VChR1 with those

of ChR1 keeps the λ_{max}, but raises the photocurrent and expression level to a satisfying level.[94] Yet another variant "ReaChR", which is based on a heavily modified combination of VChR1 and VChR2 (only helix 6 is VChR2), is activatable with light of 590 – 630 nm.[100] By swapping the last two helices of ChR1 with those of ChR2, the expression level rose as well, and yielded crystals of sufficient quality for X-ray structure analysis, which certainly is a milestone in the research of channelrhodopsins.[38] While the achievements of these chimeras have been great, the reason, why they work so much better than their native counterparts is hardly understood at all. Rather, the progress has been achieved in a combinatorial way, by systematically creating all possible combinations of helices and screening the results. Further computational studies may provide a better guide based on mechanistic insights. However, it remains doubtful, if computational models for another chimera can be generated with sufficient quality. Clearly, further X-ray structures of the ChRs would prove highly beneficial.

The aim of our analysis of the color tuning in binding pocket models of BR was to achieve an understanding of the active sites of the ChRs. We showed the sequence dependence of the color tuning and thus, the limits of the approach. The position of the negatively charged glutamate residues – helix or loop – and hence their influence on the color tuning is highly dependent on the sequence alignment to the chosen template protein. In the specific alignment underlying our studies, the additional charged residues do not cause any significant excitation energy shift, instead, differences in the binding pocket are responsible for a large portion of the shift, as shown in Table 5.7. Interestingly, the perturbation analysis of the ChR1/ChR2 seem to support that, since the most change besides the two counterions occurs at Lys 132, which causes a redshift of about -0.4 eV.[237] The glycine mutant Lys132Gly indeed causes a blueshift of 0.1 eV.[213] Therefore, large shifts of the λ_{max} for charged residues are not observed for the *true* sequence alignment as well.

It has been proposed that the major role of the electrostatic interaction with the binding pocket residues is to counterbalance the blueshift of the counterions, named "counterion quenching" by Tomasello et al.[241] The binding pocket residues have to have an overall redshift for the ultrafast photoisomerization of the retinal to occur. Since all rhodopsins contain the same cofactor and the counterions seem to shift the excitation energy to similar extents, [214,241] the counterion quenching suggests that the limits of (electrostatic) spectral tuning have to be overcome by the twisting of the retinal in strongly blue-shifted rhodopsin variants. In our analysis of the ChR binding pocket models however, we find only a slightly twisted retinal, the excitation energy of the isolated chromophore being only slightly shifted by 0.1 eV compared to BR, as shown in Table 5.5. This has also been shown for the ChR1/ChR2 X-ray structure.[213] On the other hand, we find no large redshift of the binding pocket residues in these models (Table 5.6). In the chimeric ChR1/ChR2, the dominating electrostatic effect is a large blueshift by the two counterions Glu 162 (0.87 eV) and Asp 292 (0.65 eV) and a large redshift by Lys 132 (-0.43 eV), influences of other amino acids are at 0.1 eV or below.[237] It should be noted, that due to the different alignment, we did not consider the amino acid Lys 132. Based on our analysis, which does not show a large difference between the binding pocket and a full homology model (Table 5.7), it is possible, that the effects of Lys 132 and the negatively charged glutamates, such as Glu 121, Glu 122, Glu 129 cancel each other. Further analysis clearly is required. Aside from helices 1 and 2, several replacements occur on helices 3 to 7 compared to BR, which lead to blueshifts (Table 5.2). We therefore conclude, that due to the type of the mutations in the binding pocket, our binding pocket models show blueshifts compared to BR. This is a situation not unlike sRII, which also shows an "unusually" large blueshift with respect to BR,[214] although the retinal structures are very similar,[158,221]

and the differences in the counterions contribute only a third of the spectral shift.[158] How this affects the isomerization pathway remains an unanswered question and should stimulate further studies.

The channelrhodopsin found in the green algae *Mesostigma viride*, MChR1, is the most red-shifted of the known ChRs.[61] While we did not consider MChR1 explicitly in this study, we may extend some conclusions from the study of the chlamydomonas-type ChRs to MChR1. First, several of the glutamates of the N-terminal region of the ChRs are replaced by neutral residues in MChR1 (in ChR1 numbering: Glu87Leu, Glu122Val, Glu136Ala). Overall, this should cause a redshift with respect to ChR1. Thus, the redshift of MChR1 with respect to the other ChRs may be explained by the lower number of negatively charged amino acids in the N-terminal region. However, since the homology of the MChR1 sequence to the other ChRs is lower than between any two of them,[61] a larger difference in the binding pocket composition occurs, a discussion of which is beyond the scope of the present study and certainly deserves future attention.

We modeled the color tuning in the ChRs based on multiple mutations on a BR template. A medium number of mutations in the binding pocket indeed leads to blueshifts of the excitation energy, that are of the same order as the experimentally observed absorption maxima. The large blueshift can be traced back to two origins. While the effects of any given single mutation on the BR starting structure is small, the combination of multiple mutations in the binding pocket covers the majority of the hypsochromic shift. The other significant contribution is a stronger interaction of the retinal with its counterions, although the interaction mode is more flexible than in BR, switching between a direct interaction of the Schiff base with a counterion and a water molecule. If this switching actually leads to distinct conformations, that could explain the observed spectral pattern with multiple λ_{max},[77,219] remains a challenge for further studies. Moreover, it is yet to be clarified how the multiple conformations of the retinal (about 70 % all-*trans* and 30 % 13-*cis* and even small amounts of 9-*cis*) affect the absorption spectrum.[217]

CHAPTER 6

QM/MM Simulations of Vibrational Spectra in Bacteriorhodopsin and Channelrhodopsin-2

The structural changes that occur during the photocycle of channelrhodopsin-2, especially within the hydrogen-bonded network in the active site, are important for the mechanism of the ion channel. The residues Asp 156 and Cys 128 have been found to be involved in the channel's closing mechanism, since their mutation lead to an extension of the channel's open state by up to 5 orders of magnitude. However, interpretation of these experiments was only based on the basis of homology models, which is an approach of limited value as the previous chapter showed, due to the limited sequence similarity with other microbial rhodopsins. Since the conduction of the spectroscopic experiments, the X-ray structure and more sophisticated computational models became available. In our simulations, we provide a better model, which not only explains the spectroscopic data, but also matches extensive MD simulations and the X-ray structure, which previous interpretations do not.

Parts of this chapter are published as "QM/MM simulations of vibrational spectra of bacteriorhodopsin and channelrhodopsin-2" in the journal *Physical Chemistry – Chemical Physics*.[242]

6.1 Vibrational Spectroscopy in Proteins

Vibrational spectroscopy is a showcase example of simulations and experiments going hand in hand to examine structure, dynamics and function of biomolecules. The spectroscopic experiments provide necessary information for the improvement of simulations, e.g.protonation states of titrable amino acids, conformation of co-factors, etc. while the interpretation of these experiments need models as a connection between observation and structural foundation. This approach has proven especially fruitful in the case of retinal proteins, since changes in the protein structure along the functional cycle of these proteins can be captured elegantly with difference or time-resolved FTIR experiments, as exemplified by the work on bacteriorhodopsin.[243–245] This technique is especially suited to detect changes of protonation states of glutamatic or aspartic acids, since first, the position of the ν_{COOH} band (1710 – 1760 cm^{-1}) is generally free from overlap with other residues and is well distinguishable from the symmetric and antisymmetric bands of the unprotonated species (1400 – 1570 cm^{-1}). Second, the ν_{COOH} band has an intensity among the highest in the mid-IR region, and third, the position and

shape of the ν_{COOH} band reacts sensitively to changes in the local environment, such as hydrogen-bonding interactions or the local electric field, which acts to stabilize of destabilize the C=O electric dipole moment.[246–248]

In the photocycle of BR (Fig. 1.3,p 7), the first step involves isomerization of the retinal from all-*trans* to 13-*cis*. The first intermediate, "K", can be trapped at low temperatures and its structure can be resolved by X-ray experiments (PDB code: 1m0k).[20,21] The difference between the dark state and the K intermediate can be characterized by FTIR spectroscopy,[249] which showed that Asp 115 undergoes changes in its hydrogen-bonded network. Due to BR being studied intensively in both theory and experiments, it is often used as reference system for QM/MM simulations, also of the channelrhopsins.[28,31,160,230,250,251]

In ChR2, a series of spectroscopic experiments led to the discoveries of a pair of residues, Asp 156 on helix 4 and Cys 128 on helix 3, which is involved in the channeling mechanism.[74,75,77,234,252] Replacement of each of these residues led to 100 to 10000-fold longer open states, which can be desirable for optogenetic applications.[253] These experiments were interpreted on the basis of homology models, using either sensory rhodopsin II (SRII)[234] or BR as templates,[75] since at that time a crystal structure or a more sophisticated computational model was not available. Therefore, the same issues as for the color tuning study in the previous chapter applies. Based the homology of ChR2 to other microbial rhodopsins, a direct hydrogen bond between Asp 156 and Cys 128 was proposed, which was termed "DC gate".[234] The thought is intriguing, since the homology is rather high for helices 3 and 4, and in BR, the homologous residues Asp 115 and Thr 90 indeed form a direct hydrogen bond.

However, the direct model for the DC gate does not fit to some experimental observations. First, in the X-ray structure of the ChR1/ChR2 hybrid, the side chain of Cys 128 is oriented away from Asp 156, so that the distance between the sulfur atom of Cys 128 to the nearest oxygen atom of Asp 156 is as large as 4.4 Å. This has also been confirmed in MD simulations based on a homology model of ChR2.[64] Next, our simulations show, that there is a strongly bonded water molecule between Asp 156 and Cys 128, which stays stable during extended MD simulations.[64] MD simulations of the chimera show no stable conformations of the aspartate and the cysteine, and no stable hydrogen bonds between them.[237] Finally, in the FTIR difference spectrum between the ground state and the first photointermediate P500, a symmetric band at 1735(−)/1742(+) cm^{-1},[a] was assigned to changes in the hydrogen-bonding of Asp 156, although the intensity of the band suggested the involvement of other residues or proton transfer events as well.[234,252] Hydrogen bonds involving sulfur normally are weaker than their oxygen counterparts, so that frequencies for -SH bonded carboxylic acids are expected in the range of (1745 – 1760 cm^{-1}).[254] In the case of the direct DC gate model, 1735 cm^{-1} would be an example of an exeptionally strong hydrogen bond with sulfur, stronger than the -OH...O hydrogen bond observed for the Cys128Thr mutant of ChR2 (1740(+)/1747(−) cm^{-1}) and of similar strength as observed in the difference spectrum of the K intermediate of BR with respect to wild type (1733(+)/1740(−) cm^{-1}).[249] Based on our computational model of ChR2,[37,64] which provides atomic resolution and agrees well with the crystal structure of the ChR1/ChR2 hybrid, we propose that Asp 156 and Cys 128 are bonded via a water molecule instead of a direct connection.

While DFTB has been applied successfully in vibrational spectroscopic studies in hydrogen-bonded systems, both on its own[255] and in a QM/MM setting before,[31,215,230] the effect of hydrogen bond-

[a] A negative band always corresponds to the dark state, while a positive band indicates a band in the intermediate.

ing on the ν_{COOH} band was overestimated with DFTB2 compared to full DFT, especially in strongly hydrogen-bonded systems, e.g. an acetic acid dimer.[31,215] Furthermore, despite the improvements of DFTB3 over DFTB2 with respect to hydrogen bonding energies, proton affinities and proton transfer barriers,[126] due to shortcomings of the old "mio" parameters in the description of hydrogen bonds, the performance of the DFTB3/mio combination is inferior to DFTB2 in terms of vibrational frequencies. This study on C=O stretch vibrations in microbial rhodopsins therefore is a magnificent test for the re-parametrization of DFTB for DFTB3, called "3ob".

This study pursues three goals. First, hydrogen-bonded clusters of methanol and propionic acid are analyzed via normal mode analysis to estimate the effect of hydrogen bonding on the ν_{COOH} band and to test, whether the DFTB3/3ob combination can describe this effect to reasonable accuracy. Second, the K intermediate and dark-adapted states of BR are analyzed via normal mode analysis as well to test, whether DFTB is able to recover the experimentally observed redshift from BR to K, which differ only slightly in terms of their structure. Third, the same approach is applied to the ground state of ChR2 and to the mutant Cys128Thr, to show, that the water-bridged DC gate model from our simulations agrees better with experimental data than the previous model.

6.2 Simulation Setup

6.2.1 Calculations on Gas-phase Models

All models for the gas phase calculations were built with Avogadro.[256] Normal mode analyses were carried out after geometry optimization with DFTB2/mio,[122] DFTB3/mio[126], or DFTB3/3ob.[257] A special parametrization for vibrational frequencies, "3ob-f",[257] was tested as well. These calculations were done with a stand-alone implementation of DFTB.

Density functional calculations were performed with ORCA[228] using the PBE exchange correlation functional[118] and the def2-TZVP basis and coulomb fitting basis sets.[258,259] Hybrid DFT calculations were performed with PBE0,[260] applying the RIJCOSX approximation[261] in combination with a suitable auxiliary basis.[262] Double hybrid calculations were carried out with the B2PLYP functional[263] with the invocation of the "resolution of the identity" (RI) approximation on the MP2 part[264,265] with suitable auxiliary basis set[266] and the RIJCOSX approximation on the SCF part. All computed frequencies reported in this work are unscaled harmonic frequencies.

6.2.2 QM/MM Calculations on Bacteriorhodopsin

All QM/MM calculations were carried out with DFTB/CHARMM (version c37b1)[143,155] with the mio or 3ob parameter set for the QM region, and the CHARMM22 force field for the rest of the protein.[138,139] The structure for the dark state was prepared as in the previous chapter,[160] based on an X-ray structure of 1.55 Å resolution (PDB code: 1c3w).[14] The K intermediate was modeled accordingly, based on the crystal structure of 1.43 Å resolution (PDB code: 1m0k).[20] Standard protonation states were assumed for all titrable residues, except for Asp 115, Asp 96 and Glu 204, which are known to be protonated in both states.[29,222] No restraints were applied during geometry optimization. The QM/MM boundary was chosen to be between the C_α and C_β atoms, with a standard link atom approach. The "divided frontier" scheme was applied for the charges of the MM boundary atoms.[156]

The normal mode analyses were carried out with the VIBRAN module of CHARMM in a reduced basis, consisting of a single aspartic acid side chain (Asp 115 in BR, Asp 156 in ChR2).

6.2.3 QM/MM Calculations on Channelrhodopsin-2

As a model for channelrhodopsin-2, a snapshot obtained from a MD trajectory of a previous study was used.[37,64] These simulations are based on a homology model of ChR2 with the x-ray structure of the ChR1/ChR2 hybrid[38] as template structure. For the vibrational analysis, only a monomer of the protein and all water molecules within 3 Å of any atom of that monomer were considered. Asp 156, Glu 90 and His 201 were modeled protonated.[64] Since the starting structure is a snapshot obtained from classical MD simulations and the active site is even more polar than in BR, more residues need to be included in the QM region. Here, the retinal and Lys 257, Glu 162, Asp 252, Lys 58, and 5 closest water molecules in the active site were used in all geometry optimizations on ChR2 and the Cys128Thr mutant in addition to the residues of the DC gate Asp 156 and Cys 128/Thr 128, so that the QM region for ChR2 amounts to 129 atoms. The further procedure and setup was the same as for BR. The Cys128Thr mutant was created by simply changing the residue from cysteine to threonine and re-optimizing the structure.

6.3 Computing the C=O Stretch Vibration in the Gas Phase

To be able to judge, whether DFTB is able to correctly describe the C=O stretch vibration and the effects of hydrogen bonding upon it, it needs to be compared to a proper reference in the gas phase before it can be applied usefully in proteins.

6.3.1 Choosing a Reference from Vibrational Calculations with DFTB and DFT on Small Molecules in the Gas-phase

At first, we compare the performance of DFTB to DFT exchange correlation functionals of GGA, meta-GGA, hybrid-GGA, meta-hybrid GGA and double hybrid level to chose a proper reference to compare DFTB to.

The computed harmonic normal modes deviate from the experimental fundamental frequencies in a systematic fashion. Pure DFT exchange correlation functionals of GGA or meta-GGA type underestimate the ν_{COOH} frequency, with PBE, BP86 and TPSS showing very similar behavior with a mean absolute error of 12 to 18 cm^{-1}, while the similar mPWLYP and BLYP functionals show a larger deviation of 30 cm^{-1}. Hybrid functionals overestimate the experimental frequencies of the four molecules systematically, with a mean absolute error of 50 to 86 cm^{-1}. The double hybrid functional B2PLYP performs similar to pure DFT functionals, while the other tested hybrid functional, mPW2PLYP, behaves similar to B3LYP. DFTB2, DFTB3/mio, and DFTB3/3ob have similar mean absolute errors, which are only marginally higher than the GGA functionals BLYP and mPWLYP. Compared to the previous DFTB versions, DFTB3/3ob shows the smallest systematic deviation from the experimental frequencies with a mean error of just 12 cm^{-1}. A special parametrization for stretch frequencies "3ob-f"[257] results in a systematic error of -58 cm^{-1} with respect to experiments. It indeed improves the C=O stretch frequency of H_2CO, as reported previously,[257] but is not able to improve the values for the car-

Table 6.1 Vibrational harmonic frequencies for the ν_{COOH} band of small molecules (in cm^{-1}). ME = mean error, MAE = mean absolute error, RMS = root mean squared error.

	H_2CO	HCOOH	CH_3COOH	CH_3CH_2COOH	ME	MAE	RMS
expt.	1746.0 [267]	1770.0 [267]	1779.2 [268]	1776.1 [268]			
DFTB2/mio [122]	1825.6	1772.7	1775.4	1769.6	+30.5	35.7	64.9
DFTB3/mio [126]	1869.0	1764.4	1775.8	1768.4	+26.5	34.9	61.7
DFTB3/3ob [257]	1842.9	1767.2	1758.5	1751.5	+12.2	36.2	51.0
DFTB3/3ob-f [257]	1759.8	1692.3	1696.8	1689.8	−58.3	65.2	71.7
PBE [118]	1767.8	1766.8	1766.1	1758.8	−3.0	13.9	15.5
mPWLYP [120,269]	1744.9	1741.0	1740.6	1733.1	−27.9	27.9	32.3
BLYP [119,120]	1743.0	1739.2	1739.8	1731.2	−29.6	29.6	33.7
BP86 [119,270,271]	1761.9	1759.7	1759.2	1752.2	−9.6	17.5	18.2
TPSS [272]	1770.5	1768.3	1769.8	1761.0	−0.2	12.4	14.9
PBE0 [260]	1857.6	1851.6	1848.9	1848.3	+83.8	83.8	85.4
mPW1LYP [120,269]	1835.9	1831.3	1835.8	1825.1	+64.2	64.2	66.1
B3LYP [121]	1822.4	1817.9	1817.9	1810.5	+49.8	49.8	52.4
TPSS0 [273]	1859.1	1850.3	1857.2	1850.4	+86.4	86.4	87.8
B2PLYP [263]	1791.3	1804.1	1812.9	1806.0	+35.8	35.8	36.2
mPW2PLYP [274]	1811.1	1823.0	1831.6	1824.6	+54.8	54.8	55.1

boxylic acids. DFTB2/mio, DFTB3/mio and DFTB3/3ob significantly deviate from experiment and DFT for H_2CO, so that the judgement of performance cannot be extended beyond the tested carboxylic acids. Thus we conclude that the C=O stretch vibration is already well described by the standard 3ob parameter set, thus a special parametrization is not required.

6.3.2 Shifts of the C=O Stretch Frequency due to Hydrogen-Bonding

To estimate the performance of DFT with respect to the effect of hydrogen bonding on the ν_{COOH} band, we evaluate the shift of the frequency due to one hydrogen bond with methanol and due to two hydrogen bonds in the propionic acid dimer (Table 6.2, Fig. 6.1).

The effect of one hydrogen bond (of the propionic acid's hydroxyl group to the methanol oxygen atom) is described consistently with all chosen functionals for the given basis set. The differences are more pronounced in the case of a propionic acid dimer, which may be regarded as an example of very strong hydrogen bonding. Here, DFTB3/3ob, PBE and BP86 overestimate the shift by 15 cm^{-1}. DFTB3/mio severly overestimates this shift, which is the reason for the recommendation against its usage in previous studies. The overestimation with respect to experiment is reduced for BLYP and mPWLYP as well as on the meta-GGA and hybrid functional level and vanishes almost completely on the double hybrid level of theory. We conclude, that DFTB3/3ob is able to describe the effects of hydrogen-bonding on the ν_{COOH} band as reliably as GGA-DFT for all tested systems, while for very strongly hydrogen bonding situation, the description gets improved using hybrid or even double-hybrid functionals.

Table 6.2 The ν_{COOH} frequency (in cm^{-1}) computed with DFTB and various DFT exchange correlation functionals for propionic acid (same as Table 6.1), a model of propionic acid and methanol (model 2) and the propionic acid dimer (model 14).

	CH_3CH_2COOH	model 2	model 14
expt.[268]	1776.1		1721.5
DFTB2/mio	1769.6	1727.0	1687.6
DFTB3/mio	1768.4	1716.5	1667.6
DFTB3/3ob	1751.5	1721.4	1682.0
DFTB3/3ob-f	1689.8	1657.6	1620.2
PBE	1758.8	1727.2	1683.7
mPWLYP	1733.1	1701.7	1665.8
BLYP	1731.2	1699.3	1666.2
BP86	1752.2	1719.3	1676.5
TPSS	1761.0	1729.0	1695.2
PBE0	1848.3	1817.7	1781.4
mPW1LYP	1825.1	1794.4	1760.3
B3LYP	1810.5	1780.1	1747.0
TPSS0	1850.4	1820.6	1785.1
B2PLYP	1806.0	1777.0	1750.5
mPW2PLYP	1824.6	1794.3	1768.4

6.3.3 Effect of Basis Set Choice on the C=O Stretch Frequency

Besides the chosen exchange correlation functional, the basis set naturally influences the performance of the calculation, both in terms of accuracy and computational cost. We tested the Karlsruhe basis sets: def2-SVP[258], def2-TZVP[258], def2-QZVP[275] and Dunning's cc-pVXZ (X=D,T,Q) basis sets[276] both without and augmented with diffuse functions.[277] From this small test (Fig. 6.2), we conclude that increasing the basis set size beyond triple-ζ quality does not improve the quality of the ν_{COOH} frequencies. Because def2-TZVP is more efficient than cc-pVTZ, we preferred the former for our DFT calculations.

6.4 Effect of Hydrogen Bonds on the C=O Stretch Vibration in the Gas Phase

At first, we calculated the ν_{COOH} frequencies of several hydrogen-bonded clusters of propionic acid and methanol (Fig. 6.3) by means of normal mode analysis. By comparing the frequencies of the clusters to the frequency of the free propionic acid (model 1), the resulting shift corresponds to the effect of one or several hydrogen bonds on the ν_{COOH} band. This could be regarded as simple models for the interaction of aspartic or glutamic acid side chains with serine, threonine or water molecules, although the results are not directly transferable into the protein, since steric influences may prevent optimal hydrogen-bonding geometries, and the electrostatic environment may additionally influence the C=O dipole and therefore the ν_{COOH} band. We have chosen propionic acid as a model for the aspartic acid side chain, since its spectroscopic behavior, like the solvent shifts, fits better to side chains in proteins and model peptides than acetic acid.[248]

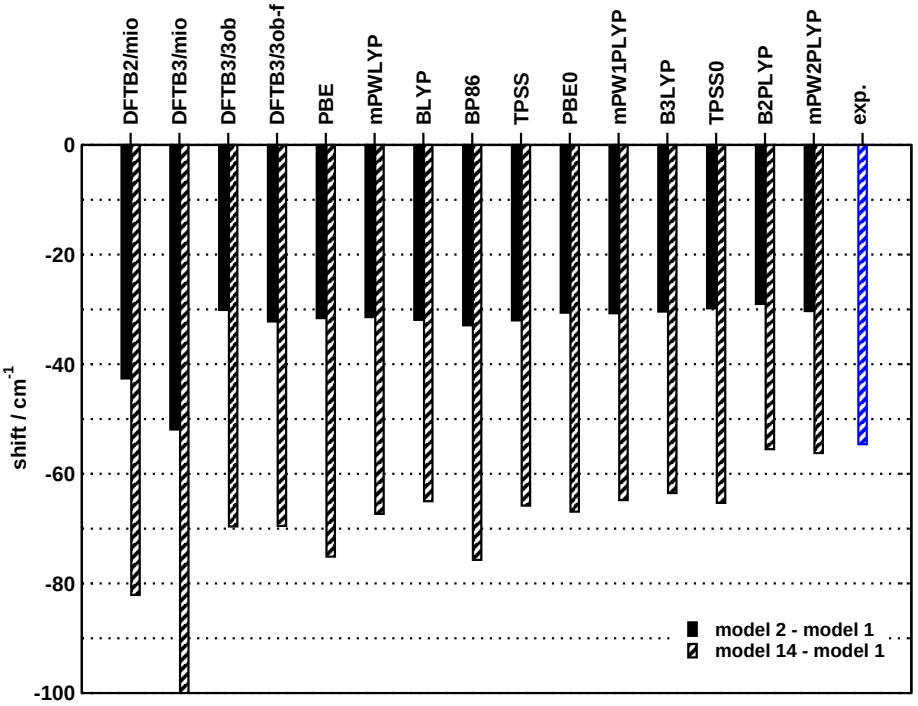


Figure 6.1 Effect of hydrogen bonds on the C=O stretch frequency, calculated as the shift of model 2 (Fig. 6.3) and propionic acid (black bars), or shift between propinic acid dimer and monomer (black/white) bars. The experimental value refers to shift between dimer and monomer.

To judge the performance of DFTB with respect to both absolute values for ν_{COOH} and shifts of the ν_{COOH} band, a proper reference method needs to be chosen. Since DFTB is derived from DFT, a pure DFT exchange correlation functional is the natural choice. The popular functionals of the GGA type compute harmonic frequencies to similar accuracy, and are overall closer to the experimental frequencies than hybrid GGA functionals.[278] The deviation from experimental data is of systematic character and can be greatly reduced by the use of functional and basis set dependent scaling factors.[279,280] In terms of the shift of the ν_{COOH} band, the hydrogen bonding needs to be accounted for properly. Many studies have evaluated the performance of DFT functionals with respect to hydrogen bonding with varying conclusions, as to which functional is "best",[278,281–284] although most of these studies focus on binding energies rather than vibrational frequencies. A general agreement appears to be, that the PBE exchange correlation functional describes hydrogen bonding interactions reliably, with a tendency for overestimation of binding energies in small water clusters with growing cluster sizes.[282] The two-center Hamiltonian matrix elements of DFTB3 are derived from PBE,[257] hence to assess the influence of the DFTB specific approximations we chose PBE to be our reference. However, to evaluate all aspects on the performance of DFT itself with respect to choice of functional, basis set truncation or anharmonicity effects deserves attention on its own. For a coarse comparison, we also show experimentally measured frequencies for the propionic acid monomer and dimer in Table 6.3, although a direct comparison between computed harmonic normal modes and experimental fundamental frequencies is not possible. Nevertheless, we note, that both PBE and DFTB3/3ob overestimate the shift between dimer and monomer by 20 cm^{-1} (Fig. 6.4). Other pure density and hybrid function-

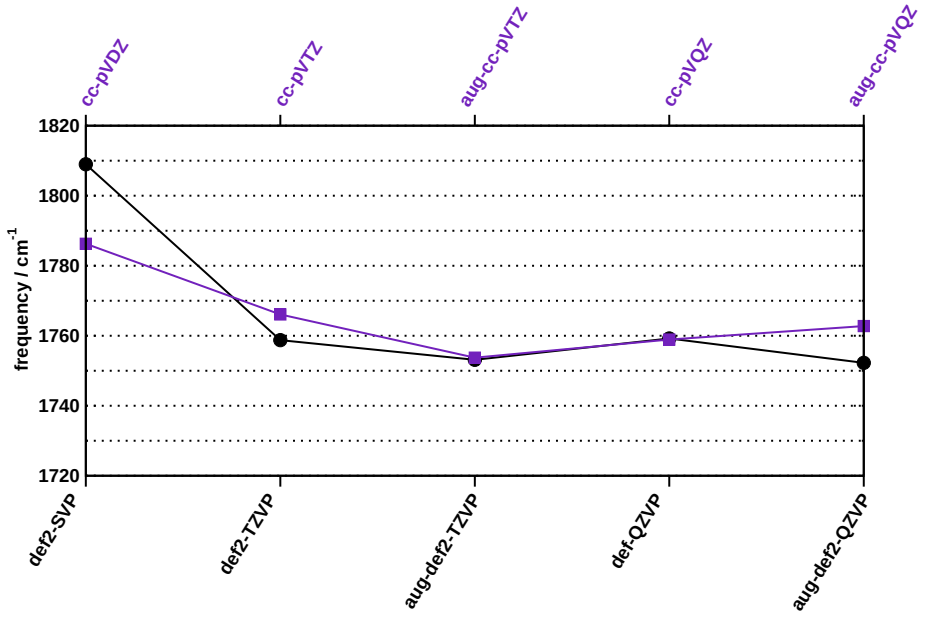


Figure 6.2 Effect of basis set choice on the ν_{COOH} frequency of propionic acid computed with PBE.

als also show that behavior. A clear improvement is made on the double hybrid functional level with B2PLYP, a functional, which yields both harmonic and anharmonic frequencies in good agreement to experiment and higher-level ab initio methods.[285]

In terms of absolute frequencies of the ν_{COOH} band in hydrogen bonded carboxylic acids (Table 6.3), the new parametrization improves the performance of DFTB3 significantly, reducing the mean absolute error from 17 cm^{-1} in DFTB2 to just 6 cm^{-1} in DFTB3. The combination of DFTB3 with the mio parameters, as it was used previously for QM/MM geometry optimizations and dynamics simulations, shows a mean absolute error of 11 cm^{-1} and is thus better than DFTB2 and worse than DFTB3/3ob. The special 3ob parameters for frequencies "3ob-f", perform inferior to DFTB3/3ob for carboxylic acids. However, for molecules of different chemical character, such as H_2CO, DFTB3/3ob deviates significantly from DFT and experiment, which shows, that the performance of semi-empirical methods needs to be evaluated carefully.[257]

For the comparison of FTIR spectra in proteins, the effects of hydrogen bonds on the ν_{COOH} band rather than absolute frequencies need to be described correctly. It was shown in earlier work, that DFTB2 overestimates the effect of hydrogen bonds compared to DFT for strongly hydrogen-bonded systems, such as the acetic acid dimer.[31,215] This effect is even more pronounced for the combination of DFTB3 with the mio parameter set, especially for the propionic acid dimer (model 14). This overestimation with respect to DFT is removed with the new parameter set (Fig. 6.5). If the propionic acid forms one hydrogen bond with the OH group as donor (model 2) or the C=O as acceptor (model 5), the frequency of the ν_{COOH} band is red-shifted by 30 cm^{-1}. Thus, by means of the frequency only, these two situations can not be distinguished. Model 3 shows a larger redshift of ca. 60 cm^{-1}, but its configuration corresponds to one hydrogen bond to the hydroxyl group and one distorted interaction with the carbonyl group. If the OH group acts as acceptor (model 4), ν_{COOH} gets blue-shifted by 15 cm^{-1}. These results agree well with a similar study on butanoic acid/methanol clusters with hybrid

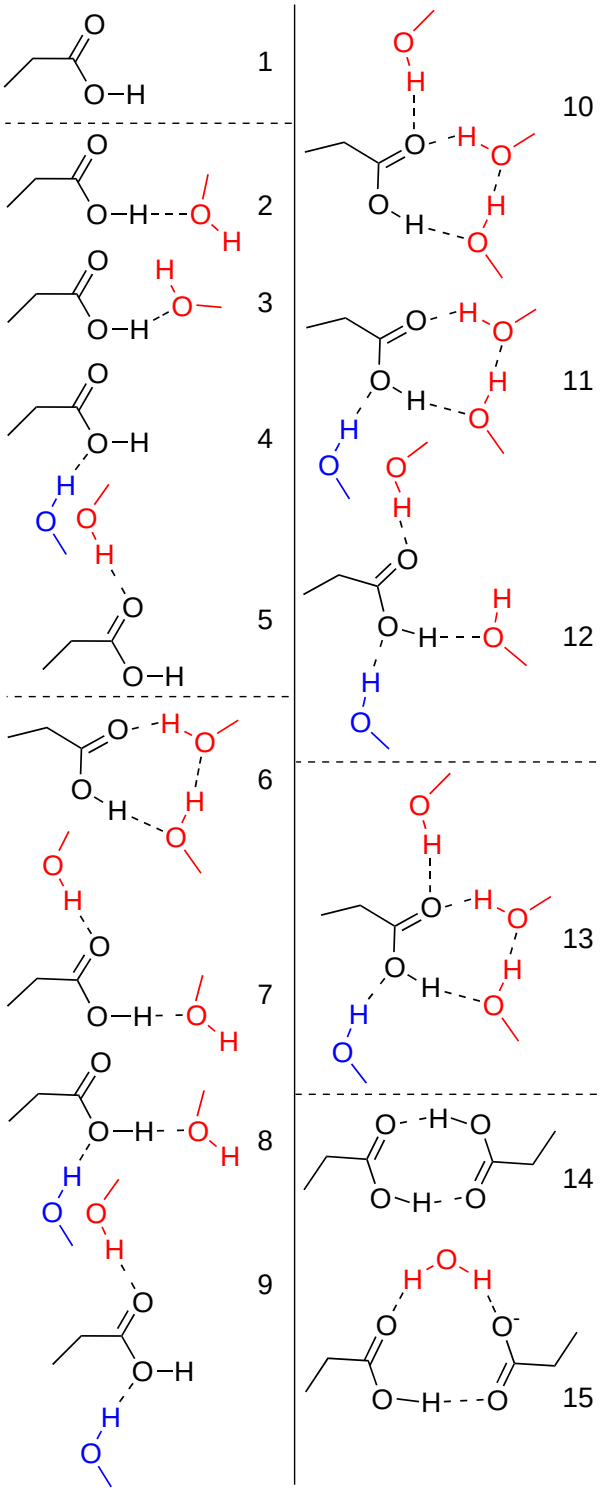


Figure 6.3 Gas-phase models of propionic acid and methanol, hydrogen-bonded in different conformations. Molecules in red lower the C=O stretch frequency, while molecules in blue cause blueshifts.

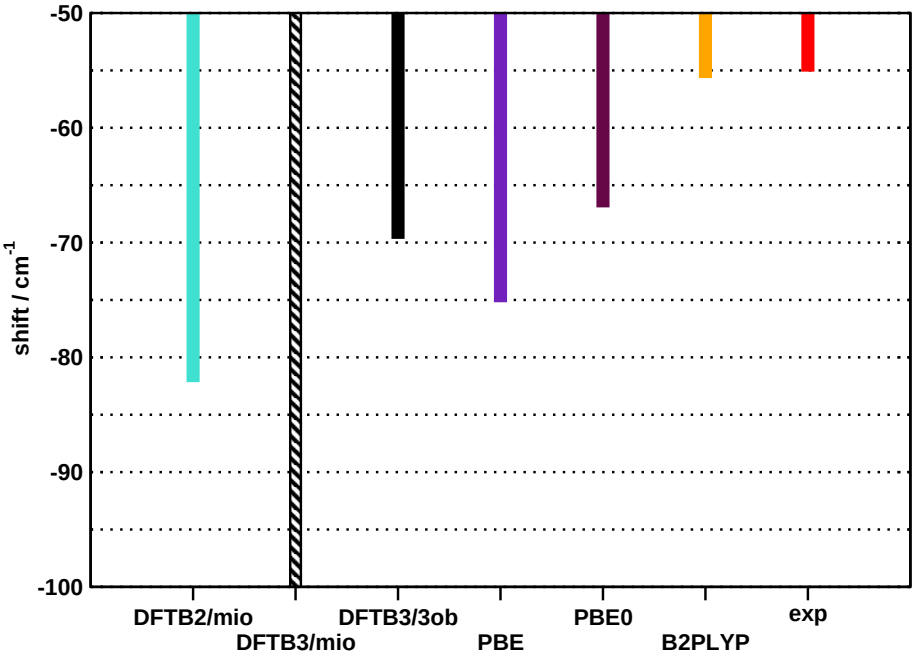


Figure 6.4 Shifts of the ν_{COOH} frequencies of the propionic acid dimer with respect to the monomer, computed with DFTB and DFT and compared to experiment.[268]

Table 6.3 The ν_{COOH} band of hydrogen-bonded clusters of propionic acid and methanol. In models 14 and 15, the strongly dipole-allowed antisymmetric mode is shown. DFT calculations with def2-TZVP basis set. Numbers given in cm^{-1}.

model	H-bonds	DFTB2/mio	DFTB3/mio	DFTB3/3ob	PBE	PBE0	B2PLYP	expt.[268]
1	0	1770	1768	1752	1759	1848	1806	1776
2	1	1727	1716	1721	1727			
3		1704	1688	1696	1699			
4		1788	1797	1767	1772			
5		1736	1719	1715	1724			
6	2	1686	1665	1675	1669			
7		1694	1671	1678	1701			
8		1750	1750	1738	1740			
9		1758	1753	1736	1737			
10	3	1652	1621	1639	1639			
11		1702	1684	1685	1679			
12		1714	1700	1702	1703			
13	4	1676	1652	1657	1648			
14		1688	1668	1682	1684	1786	1750	1722
15		1774	1696	1677	1697			
ME w.r.t. PBE		+16	-2	-4				
MAE		17	11	6				
RMS		25	14	9				

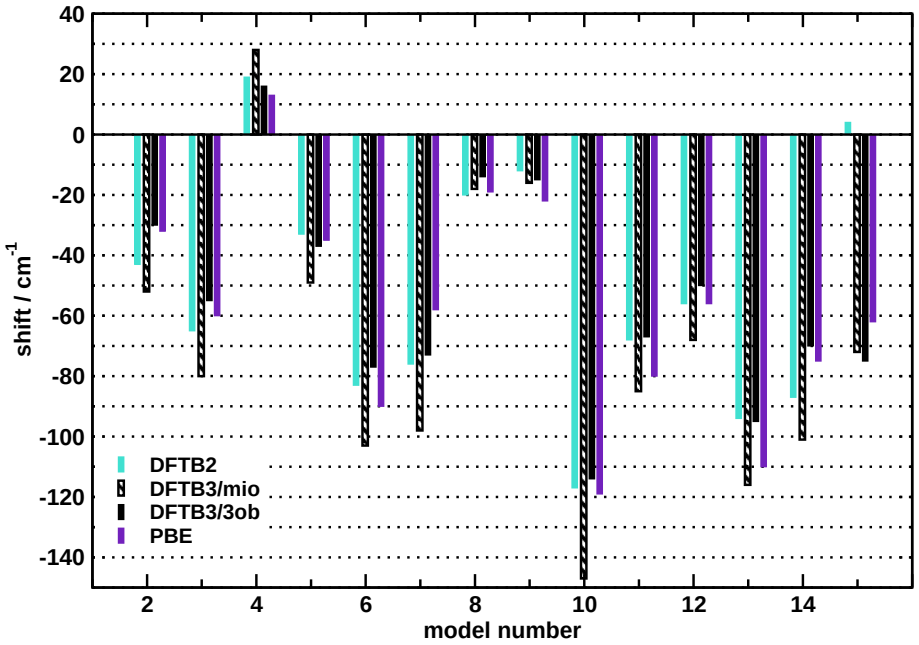


Figure 6.5 Shifts of the ν_{COOH} band of propionic acid, due to hydrogen bonds with methanol. Errors for the shifts with respect to PBE (ME/MAE/MSE): DFTB2: +6/12/20, DFTB3/mio: -12/15/19, DFTB3/3ob: +4/8/9.

density functional theory.[254]

In addition, we also examined the effect of one hydrogen bond between propionic acid and methanethiol (models 2 to 5 in Fig. 6.3, with CH_3SH instead of CH_3OH) as a model for the possible interaction between aspartic acid and cysteine in ChR2. Hydrogen bonds involving sulfur are expected to be weaker than those with oxygen, so that the shifts of the ν_{COOH} band due to this hydrogen bond are expected to be smaller. Indeed, DFTB3 describes the shifts due to CH_3SH to be 5 to 10 cm^{-1} smaller than due to CH_3OH in reasonable agreement to DFT (Fig. 6.6), although DFTB predicts a greater difference for model 2 between sulfur and oxygen as acceptor. For sulfur, the 3ob parameters are under development, so that all calculations involving sulfur are performed with the mio parameter set.

To summarize, DFTB3 in combination with the new 3ob parameter set shows a good agreement to full DFT for the vibrational analysis, both in terms of the absolute frequency, as well as the shifts due to hydrogen bonds for systems like hydrogen bonded carboxylic acids. We find, that for hydrogen-bonded carboxylic acids, the standard 3ob set performs well, although in other cases, like H_2CO, large deviations from DFT and experiment are observed. The shift of the propionic acid dimer with respect to the monomer, as an example for strong hydrogen-bonding, shows, that both DFTB and DFT overestimates the experimental shift by 20 cm^{-1}, which gets improved using hybrid or even double hybrid exchange correlation functionals.

6.5 The C=O Stretch Vibration in Bacteriorhodopsin: Dark State vs. K Intermediate

As a next test, we computed the C=O stretch vibration of Asp 115, both in the dark state and in the K intermediate to additionally examine the effect of the environment (electrostatics) on the C=O stretch

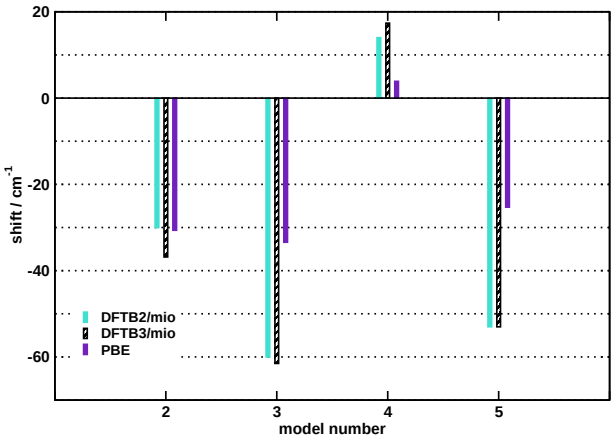


Figure 6.6 Shifts of the ν_{COOH} band of propionic acid due to hydrogen bonds with one CH_3SH molecule (models 2 – 5 of Fig. 6.3, with CH_3OH replaced by CH_3SH).

vibration. In both structures, the OH group of Asp 115 acts as a donor for a hydrogen bond with Thr 90, while it also acts as an acceptor for a hydrogen bond with water 511, a water molecule that is resolved in both x-ray structures 1c3w and 1m0k (Fig. 6.7). The hydrogen-bonding pattern of Asp 115 therefore corresponds to model 8 of the gas-phase clusters (Fig. 6.3).

Figure 6.7 Hydrogen-bonded network around Asp 115, Thr 90 and water for the QM/MM optimized structure of BR.

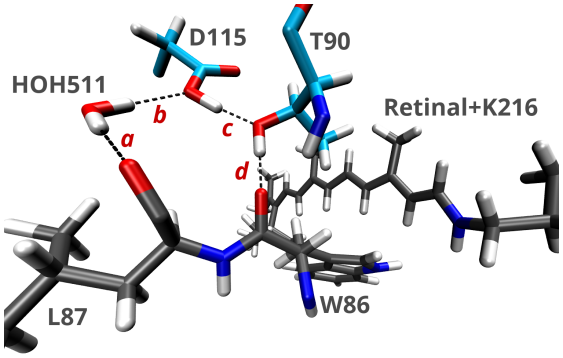


Table 6.4 O-O distances (in Å) of the hydrogen-bonded network around Asp 115 in the dark state and K intermediate of BR.

dist. between	label	O-O distances	
		BR	K
HOH 511 – Leu 87	a	2.80	2.82
HOH 511 – Asp 115	b	2.99	2.93
Asp 115 – Thr 90	c	2.59	2.61
Thr 90 – Trp 86	d	2.57	2.57

In the K intermediate, the retinal has isomerized from all-*trans* to 13-*cis*, while the rest of the protein remains in a conformation very similar to the dark-adapted state. The difference FTIR spectrum of the K intermediate vs. the dark state of BR shows a symmetric difference band at 1733(+)/1740(−) cm^{-1}, which vanishes for the Asp115Glu or Asp115Gln mutants.[249] Therefore, this feature has been assigned to a change in the ν_{COOH} band of Asp 115. We examine the ν_{COOH} band with different QM regions. In order to describe the hydrogen-bonded Asp 115 correctly, its hydrogen-bonding partners need to be included in the QM region (Asp 115+Thr 90+WAT, Table 6.5). If water 511 is deleted from the structure (Asp 115+Thr 90-WAT), the remaining shift of the ν_{COOH} band is mainly caused by Thr 90. As expected from the model calculation in the gas phase, the hydrogen bond with Thr 90 red-shifts ν_{COOH}

Table 6.5 The ν_{COOH} band of Asp 115 in the dark state and K intermediate of BR.

			ν_{COOH} / cm^{-1}		
structure	QM region	expt.[249]	DFTB2	DFTB3/mio	DFTB3/3ob
BR		1740			
	Asp 115+Thr 90+WAT		1740	1740	1732
	Asp 115+Thr 90		1749	1744	1735
	Asp 115+Thr 90-WAT*		1730	1719	1722
	Asp 115		1757	1761	1749
K		1733			
	Asp 115+Thr 90+WAT		1737	1737	1728
	Asp 115+Thr 90		1745	1741	1732
	Asp 115+Thr 90-WAT		1727	1719	1721
	Asp 115		1754	1757	1746

∗ water 511 deleted before structure optimization.

by 30 cm^{-1}. The water molecule blue-shifts the frequency by 10 cm^{-1}. The difference between the QM regions Asp 115+Thr 90+WAT and Asp 115+Thr 90 describes the polarization of water HOH511, for it is described by DFTB3 in the former and by the TIP3P force field in the latter QM region. This effect is only minor with a shift of 3 – 4 cm^{-1}.

Only these three residues need to be included in the QM region. The frequencies computed with also Lys 216(+ retinal), Asp 212, Asp 85, waters 401, 402 and 406 in the QM region are virtually indistinguishable from the values shown in Table 6.5, although minor structural differences occur in the Schiff base region of the retinal for the K state. The retinal is more twisted, if described with the CHARMM force field but since the ν_{COOH} band is determined by local effects, the impact of this structural difference is only about ± 1 cm^{-1}.

The redshift of the ν_{COOH} band between the dark state and the K intermediate is reproduced well with DFTB compared to the experiment. It is interesting to note, that both the hydrogen-bonded network of Asp 115 and the C=O bond lengths are very similar in both states. Subtle changes in the O–O distances (Table 6.4) and the different conformation of the retinal (13-*cis* in K and all-*trans* in the dark-adapted state) are able to cause a shift of 7 cm^{-1}, which is described well with DFTB.

6.6 The C=O Stretch Vibration of Asp 156 in the DC Gate of ChR2

The difference FTIR spectrum between the dark state and the first photointermediate of ChR2, P500, shows a symmetric difference band at 1735(−)/1742(+) cm^{-1}. Compared to the BR/K difference spectrum, the change is opposite in sign and larger in intensity, so that changes other than the hydrogen-bonded network of Asp 156/Cys 128 may be involved. For the Cys128Thr mutant, the difference band pattern is similar to BR (1747(−)/1740(+) cm^{-1}), but blue-shifted by 7 cm^{-1}, implying a weaker hydrogen bond. Even less is known about the P500 state than about the dark state, so that any attempt of modeling remains an issue for future studies. Here, we compute the ν_{COOH} bands of a computational model of ChR2,[64], the Cys128Thr mutant based on that model, and compare to the experiments. MD simulations and analysis with a 3D reference interaction site model (3D RISM)[64,286] suggested a much

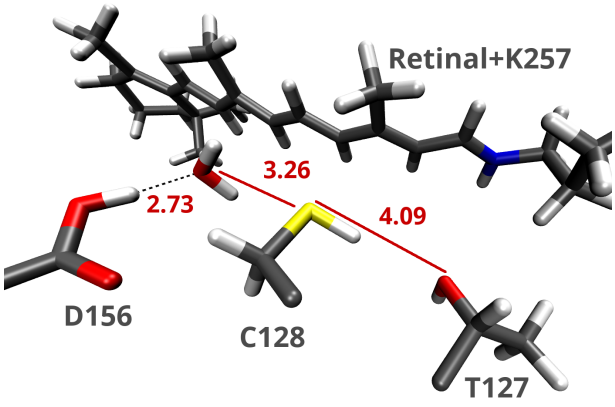


Figure 6.8 Hydrogen-bonded network around Asp 156 for a QM/MM optimized structure of ChR2. Distances are O–O and O–S distances respectively (in Å).

larger number of internal water molecules than resolved in the X-ray structure of the ChR1/ChR2 hybrid[38] or BR.[14] The analysis revealed a stable position of one water molecule between Cys 128 and Asp 156, which is not present in BR and therefore any homology models based on a BR template. In Fig. 6.8, the structure of the "indirect" model for the DC gate is displayed. Asp 156 is hydrogen-bonded to one water molecule, while the Cys 128 is oriented to the side chain oxygen atom of Thr 127, the S-O distance being 4.09 Å. The computed frequencies for the ν_{COOH} band agree well with the experimental data for ChR2, which shows, that the FTIR experiments can be interpreted on an indirect model as well, where Asp 156 is hydrogen-bonded to one water molecule.

A direct interaction between Asp 156 and Cys 128 is not observed in MD simulations, neither in ChR2,[64] nor in the ChR1/ChR2 hybrid.[237] Instead, the simulations indicate a stable water molecule bridging Cys 128 and Asp 156. Furthermore, the orientation of Cys 128 in the X-ray structure of the ChR1/ChR2 hybrid, on which our homology model is based, is unsuitable for a direct interaction with Asp 156. Therefore, a DC gate that features a direct interaction is not compatible to our homology model.

The Cys128Thr mutant looks very similar to the dark state of BR (Fig. 6.9) without any water molecule between Asp 156 and Thr 128. The ν_{COOH} band is shifted to 1747 cm^{-1} in the ground state of ChR2, which is 7 cm^{-1} higher than in BR, implying a weaker hydrogen bond between Asp 156 and Thr 128. In BR, the position of the water molecule near Asp 115 is well resolved in the X-ray structure,[14] while in Cys128Thr, no water molecule was found at that position.[64] Additionally to the model obtained from MD, we also built a model with a water molecule near the OH group of Asp 156, similar to BR (shown in green in Fig. 6.9). However, as the O–O distances are larger compared to BR (Table 6.6), the influence of the water molecule on the ν_{COOH} frequency is smaller (Table 6.7). The difference between the C=O stretch frequencies between ChR2 and the Cys128Thr mutant is 12 cm^{-1} as determined by the experiment. Our calculations on the other hand predict similar C=O stretch frequencies for ChR2 and its mutant. One reason behind this deviation is based on the setup of the structure for the mutant. We started from an optimized ChR2 structure, changed the cysteine to threonine and optimized again. Therefore, the structural relaxation is only local. Clearly, extended MD simulations would lead to a better representative structure of the Cys128Thr mutant, however, this is

Table 6.7 The ν_{COOH} band of Asp 156 in the wild type ChR2 and in the Cys128Thr mutant computed with DFTB (3ob parameter for sulfur are under development).

structure	QM region	ν_{COOH} / cm^{-1} exp.[234]	2/mio	3/mio	3/ob
ChR2		1735			
	Asp 156+Cys 128+WAT		1740	1742	—
Cys128Thr		1747			
	Asp 156+Thr 128		1738	1733	1735
	Asp 156+Thr 128+WAT		1743	1733	1737

beyond the scope of this work.

Figure 6.9 Hydrogen bonded network around Asp 156 in the Cys128Thr mutant. The water molecule (green) is not present in MD simulations, but inserted manually before structure optimization.

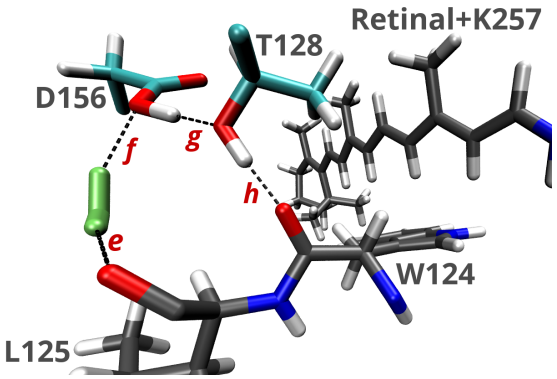


Table 6.6 O-O distances (in Å) of the hydrogen-bonded network around Asp 115 in the dark-adapted and K state of BR.

dist. between	label	O–O distances
WAT – Leu 125	e	2.94
WAT – Asp 156	f	3.08
Asp 156 – Thr 128	g	2.66
Thr 128 – Trp 124	h	2.64

Clearly, the calculated vibrational frequencies of Asp 156 in ChR2 deviate more from the FTIR experiments than in BR, where the structural uncertainty is very low. It cannot be expected, that our homology model of ChR2,[64] which is based on the X-ray structure of the ChR1/ChR2 hybrid and refined by MD simulations, possesses the same structural certainty as BR. The example of BR/K shows, that small changes, like subtle conformation changes, minor deviations in the hydrogen-bonding network or a different configuration of the chromophore, lead to shifts of ν_{COOH} band of 7 cm^{-1}. The major effect, which is the shift of the ν_{COOH} of 1779 cm^{-1} (gas phase) to 1740 cm^{-1} as in the protein, is dominated by local effects, which are described well with DFTB in the QM/MM framework.

6.7 Conclusions

We showed that DFTB3 in combination with the 3ob parameter surpasses the accuracy of previously employed versions for the vibrational analyis of carboxylic acids, with a mean absolute error of only 6 cm^{-1} with respect to PBE for the frequencies of the ν_{COOH} band, and 7 cm^{-1} for the shifts due

to hydrogen bonding interaction with CH_3OH. Nevertheless, these calculations also made clear, that the performance of semiempirical methods can not be generalized to other type of molecules. While the C=O stretch vibration in propionic acid is computed 24 cm^{-1} below the experimental value, in excellent agreement to PBE, in formaldehyde, DFTB3/3ob yields a vibrational frequency, that is almost 100 cm^{-1} above the experiment and 80 cm^{-1} above PBE.

We studied the FTIR difference spectrum between the dark state and the K intermediate of BR to show that subtle changes in the hydrogen-bonding network of Asp 115 are able to cause a redshift of the ν_{COOH} band of 7 cm^{-1}. As an approximative method, it is not the aim of DFTB to seek quantitative agreement, but rather to recover the effects of the two different structures on the ν_{COOH} frequency, which it does well.

Unfortunately, the availability of "good" models for the ChRs lagged behind the availability of experimental data for a long time, since models based on X-ray crystallographic experiments and MD simulations only appeared recently. Previously, experimental data have been interpreted often on the basis of homology models, which may or may not be suitable, depending on the choice of template and structural refinement. The DC gate in ChR2 is a structural motif, which is involved in the kinetics of the ion channel Replacements of either Asp 156, Cys 128, or both significantly prolongs the open state. Changes in the structure of the DC gate are visible in the FTIR difference spectrum which have been interpreted based on homology to other familiar microbial rhodopsins. Accordingly, a direct hydrogen bond between the thiol group of Cys 128 and the carbonyl group of Asp 156 has been postulated. However, a direct model for the DC gate is inconsistent with both X-ray crystallographic and the FTIR experiments themselves, since hydrogen-bonded carboxylic acids are usually found 15 – 20 cm^{-1} above the reported value of ChR2 (1735 cm^{-1}). An indirect model, where Asp 156 and Cys 128 interact via a water molecule, can explain the FTIR experiments and additionally is consistent to the X-ray structure of the ChR1/ChR2 hybrid. The vibrational analysis of Asp 156 based on our computational model shows a general agreement with the experimental data, with remaining uncertainties due to our ChR2 model.

The bridging water molecule is an interesting idea for the mechanism of ChR2. If the Schiff base gets reprotonated by Asp 156 in the photocycle, the water molecule might shorten the distance between Asp 156 and the retinal's nitrogen. Further, presence or absence of the water molecule might be responsible for the prolonged open times of ChR2 mutants.[64] As a further test of our proposed model, it would be interesting to perform both, simulations as well as FTIR experiments on the ChR1/ChR2 hybrid. First simulations indicate, that a stable structure for the DC gate cannot be reached without the presence of the water molecule and a direct hydrogen bond is not observed, which matches our observations on ChR2.[237] Further characterization of the DC gate requires the simulation of the first intermediate of the photocycle, P500. Currently, no structural model is available for P500. Due to the short transition time between the dark state D470 and P500 ($\approx$3 ps), the simulation of the P500 intermediate is not out of reach with current technology, so that future work will be able to directly compare the situations from BR/K and D470/P500.

CHAPTER 7

The N-H Vibration of Retinal

In the FTIR spectrum of the dark state of BR, a broad and weak negative band around 2700 – 3100 cm^{-1} is found,[287] which vanishes in the K intermediate and is shifted to 2100 – 2300 cm^{-1} in D_2O. This lead to the hypothesis, that changes in the hydrogen-bonding network of the Schiff base are responsible for this band. In BR, the Schiff base forms a hydrogen bond with a water molecule (Fig. 7.1(a)), while in the K intermediate, due to the twist of the retinal, the Schiff base has no hydrogen-bonding partner (Fig. 7.1(b)), explaining the disappearance of the weak band in K. In the experiment, O-H and N-H stretch vibrations of the protein are observed around 3300 – 3500 cm^{-1} and are hidden beneath the absorption of the bulk solvent. Band assignment can therefore only be made indirectly via experiments with deuterated agents.

QM/MM simulations are very powerful for the computation of vibrational spectra, because the total spectrum can be decomposed into contributions of different subspecies by intelligent design of the QM region. In that fashion, the contribution of the retinal and its hydrogen-bonded network can

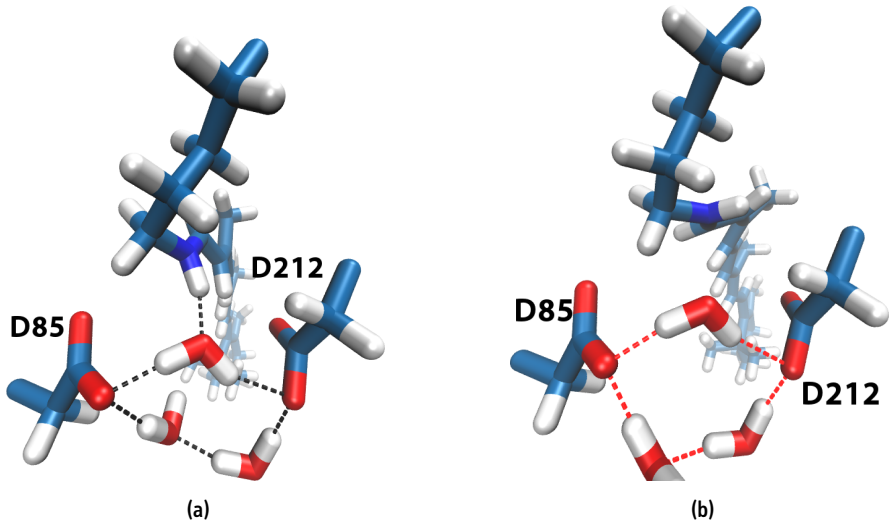


Figure 7.1 **(a)** In the dark state of BR, the Schiff base forms a hydrogen bond with a water molecule. **(a)** In the K intermediate, the hydrogen bond is disrupted due to the twist of the retinal.

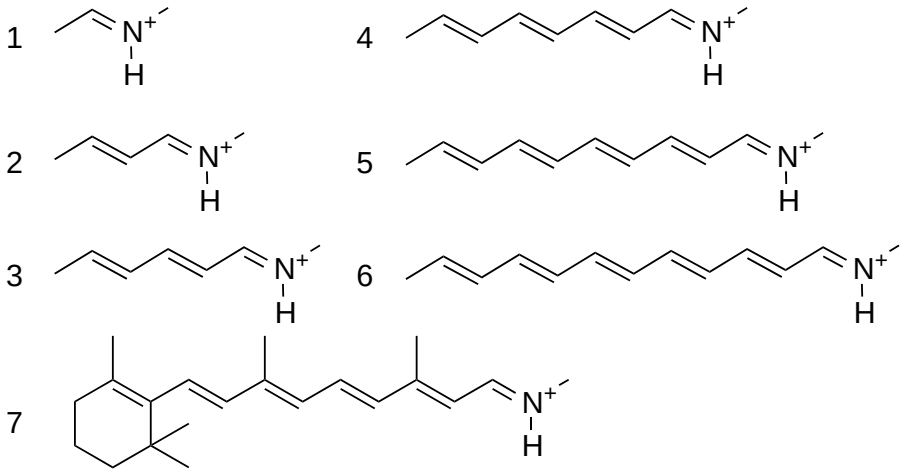


Figure 7.2 Gas-phase models for the calculation of N-H stretch frequencies.

be singled out. In contrast to the study of C=O stretch vibrations, where the peaks in the difference spectrum were sharp[249] and did not indicate any conformational contributions, the broad character of the weak band suggests, that normal mode analysis alone will not fully describe the situation. Therefore, we apply the Fourier-transformed time correlation function (FTTCF) formalism, that takes into account the dynamics of the system and allows for a more realistic simulation of vibrational spectra.

7.1 Simulation Protocol

In the gas phase, all calculations were performed as described for the C=O vibration in the previous chapter (cf. 6.2,p 71). In addition, we also perform RI-MP2[264,265] calculations with the def2-TZVP basis set and suitable auxiliary basis.[258,266] The gas phase Schiff base models are shown in Fig. 7.2. In addition to normal mode analysis, vibrational spectra were simulated using the Fourier-transformed time correlation function (FTTCF) formalism. Spectra were collected from averaging over 100 separate QM/MM MD simulations à 40 ps, with settings as described in section 5.4.1. The spectra were averaged and a simple correction factor was applied to approximate the quantum effects of nuclear motion, as described in section 3.1.2 (p 25). N-D vibrations were computed by exchanging the masses of the respective protons to the mass of deuterium and rediagonalizing the Hessian matrix. The structures were not re-optimized.

7.2 Performance of DFTB with Respect to the N-H Stretch Vibration

While the DFTB method has been applied successfully in vibrational studies, both in pure QM simulations[255,288–291] and in QM/MM settings,[31,215,230,242,292] we test the performance of DFTB2 (with mio parameters)[122] and DFTB3 (with 3ob parameters)[126,257] with respect to the N-H stretch vibration (ν_{NH}) in the retinal, and the effect of deuteration and hydrogen bonds on it. At first, we compare DFTB with DFT and MP2 (Table 7.1).

Table 7.1 The N-H stretch vibration in several model molecules in gas-phase (in cm^{-1}). Errors are given with respect to PBE and B2LYP as reference in the form ME/MAE/RMS (ME = mean error, MAE = mean absolute error, RMS = root mean squared error).

model	DFTB2	DFTB3	PBE	PBE0	B2PYLP	MP2
1	3477	3268	3403	3526	3502	3496
2	3498	3290	3437	3558	3529	3519
3	3514	3302	3454	3574	3545	3533
4	3523	3311	3465	3588	3557	3542
5	3529	3321	3473	3594	3564	3551
6	3534	3319	3479	3598	3568	3556
7	3539	3324	3483	3603	3574	3559
Δ PBE	+61/61/61	−150/150/150		+121/121/121	+92/92/92	+81/81/81
Δ B2PLYP	−31/31/32	−242/242/242	−92/92/92	+29/29/29		−11/11/11

Table 7.2 The N-D stretch vibration in several model molecules in gas-phase (in cm^{-1}). Errors are given with respect to PBE and B2LYP as reference in the form ME/MAE/RMS.

model	DFTB2	DFTB3	PBE	PBE0	B2PYLP	MP2
1	2554	2403	2495	2586	2569	2567
2	2567	2416	2518	2608	2588	2582
3	2577	2424	2530	2619	2599	2591
4	2583	2430	2538	2629	2607	2597
5	2588	2436	2543	2633	2612	2603
6	2591	2435	2548	2636	2614	2607
7	2594	2438	2550	2639	2618	2609
Δ PBE	+48/48/48	−105/105/105		+90/90/90	+69/69/69	+62/62/62
Δ B2PLYP	−21/21/21	−174/174/174	−69/69/69	+20/20/21		−7/7/7

With increasing chain lengths, the ν_{NH} frequency increases, which all methods describe well. Unfortunately, experimental gas-phase spectra of chemically similar molecules are unavailable. While there are no experimental frequencies to compare to, the systematic character of the errors indicate, that deviations can be corrected by the application of global scaling factors. Omitted from Table 7.1 is the combination DFTB3/mio. For completeness sake, it shall be noted, that the overall performance was identical to DFTB2/mio, while individual frequencies differed by 4 cm^{-1} maximally. Compared to DFTB3, the systematic offset of DFTB2 compared to the other methods is smaller.

It is known, that DFT tends to overestimate the conjugation, while MP2 overestimates the alternating single/double bond character.[207] Since MP2 and the tested exchange-correlation functionals show a similar increase in the ν_{NH} frequency, the influence of the different description of conjugation seems to be systematic es well.

The range, in which the ν_{NH} frequency appears overlaps with the O-H stretch vibration of water, so that a direct assignment is impossible in experiments. Bands can be assigned from frequency shifts after deuteration, where only vibrations, that are affacted by the H/D exchange are visible, such as N-H or O-H bands. The N-D stretch vibrations occupy a rather sparsely populated region of the spectrum, 2000 – 2700 cm^{-1} (Table 7.2).

For the deuterated molecules, the deviations of the methods with respect to each other is slightly re-

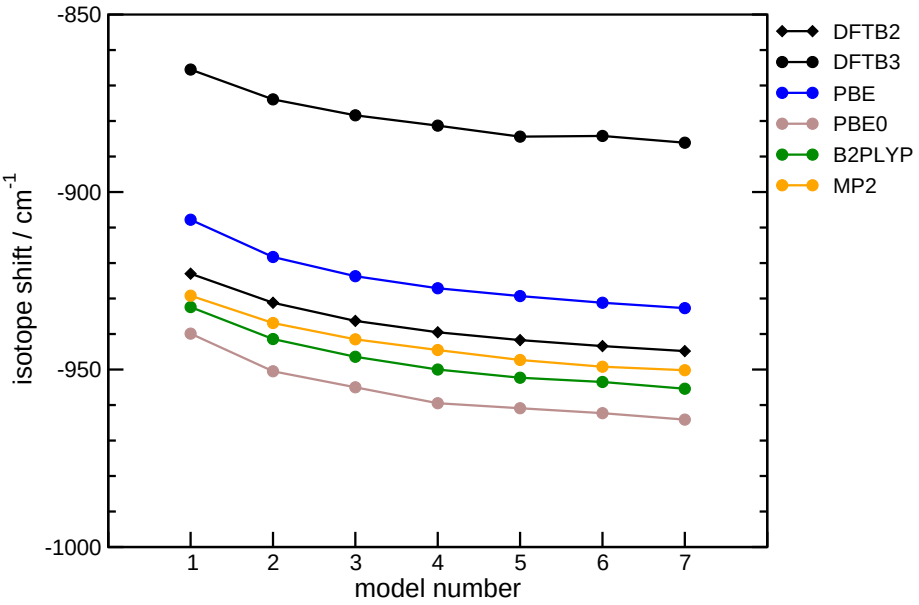


Figure 7.3 Isotope shift of the ν_{NH}-band in gas-phase models.

Table 7.3 Mean errors of isotope shifts (Fig. 7.3) with respect to PBE and B2PLYP (in cm^{-1}). The mean absolute errors and root mean squared errors are numerically identical.

reference	DFTB2	DFTB3	PBE	PBE0	B2PLYP	MP2
PBE	−13	+45		−32	−23	−18
B2PLYP	+10	+68	+23	−9		+4

duced. The errors are again of very systematic character, from which it can be concluded, that the N-H and N-D stretch vibrations are described to similar accuracy. This is also reflected in the uniform distribution of the isotope shifts (Fig. 7.3).

All methods describe the isotope shift of the ν_{NH} vibration to a similar degree, especially the agreement of DFTB2 with PBE and B2PLYP is encouraging.

In BR, the retinal is hydrogen-bonded to one water molecule. This causes an additional redshift of the ν_{NH} frequency. To test, how DFTB fares against the other methods, we re-optimized the gas-phase models with one water molecule placed below the N-H bond. In the calculations, the N-H stretch vibration contributes to multiple normal modes due to symmetric and antisymmetric combinations with C-H bond stretches. The relative intensities of these combinations are different for each method, which impedes an exact comparison. Here, we designate the normal mode with N-H vibrational contribution and highest intensity *the* ν_{NH} band.

In Table 7.4, the ν_{NH} frequencies are reported for the hydrogen-bonded molecules in the gas phase. During the optimization, the structures of model 6 computed with B2PLYP and MP2 became highly twisted. Due to software issues, the frequency analysis of model 7 could not be performed with B2PLYP and MP2. Therefore, we can only compare DFTB to PBE and PBE0. DFTB2 again has a very systematic error with respect to PBE. The deviation of DFTB3 with respect to PBE is significantly smaller. While this indicates a good performance for the absolute ν_{NH} frequency in protonated Schiff base models,

Table 7.4 The N-H stretch vibration of hydrogen-bonded model molecules in gas-phase (in cm^{-1}). Errors are given with respect to PBE as reference in the form ME/MAE/RMS.

model	DFTB2	DFTB3	PBE	PBE0	B2PLYP	MP2
1	3025/3056*	3009/3035*	2946	3106	3144	3119
2	3159	3094	3049	3202	3232	3180/3197
3	3273	3127	3122	3311	3280	3237
4	3239	3152	3161	3299	3313	3266
5	3268	3170	3190	3388	3353	3287
6	3281	3183	3211	3433	3412†	3512†
7	3304	3203	3239	3414	—‡	—‡
Δ PBE	+92/92/97	+5/31/38				

* The symmetric and antisymmetric combinations of N-H and C-H bond stretches have similar intensity and are reported both.

† optimized structures are highly twisted/bent

‡ due to software limitations, these numbers could not be obtained

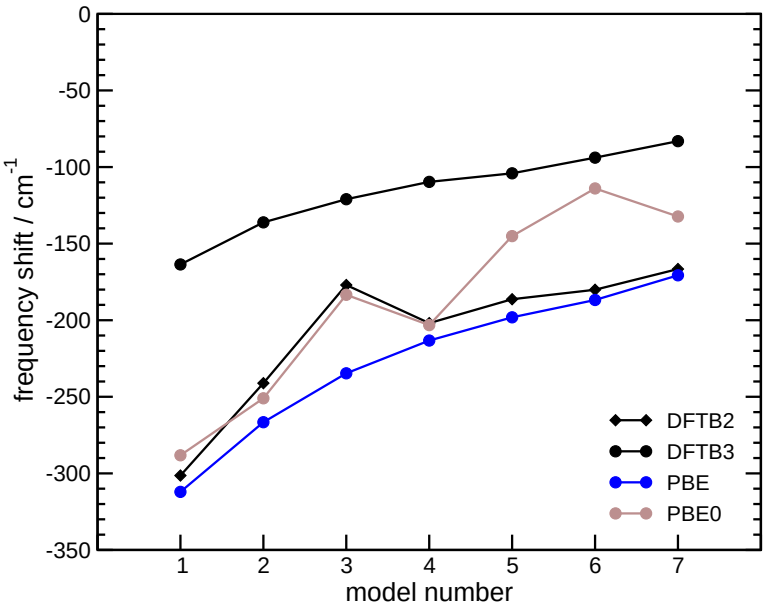


Figure 7.4 Shifts of the ν_{NH} frequency due to hydrogen-bonding in the gas phase. Errors of DFTB2 with respect to PBE (+18/18/25 cm^{-1}); errors of DFTB3 with respect to PBE (+110/110/112 cm^{-1}).

the shifts due hydrogen-bonding will not be reproduced well by DFTB3. Compared to the C=O stretch vibration, where the hydrogen bond red-shifts the ν_{CO} frequency by ≈70 cm^{-1}, the effects are larger with >100 cm^{-1} for the N-H stretch.

It is shown in Fig. 7.4, that the shifts due to hydrogen bonding are hard to capture. While both PBE and DFTB3 show a decrease of the shift with increasing chain length of the Schiff base model, PBE0 and DFTB2 show deviating behavior. Therefore, a general assessment of the performance of DFTB with respect to the hydrogen-bonding shift cannot be given. If one considers the all-*trans*-retinal separately, the deviation in the shifts of DFTB2 from PBE is only 11 cm^{-1}. Compared to PBE0,

Table 7.5 The N-D stretch vibration of hydrogen-bonded model 7 in the gas phase (in cm^{-1}).

model	DFTB2	DFTB3	PBE	PBE0
7	2428	2355	2380	2507

both seem to overestimate the effect of hydrogen-bonding. The reason, why the behavior of the tested methods is so heterogeneous is not clear. Structural differences in the optimized geometries can be ruled out. Also, the ν_{NH} band with highest intensity is the antisymmetric combination of N-H and C-H bond stretches. Therefore, different constitutions of the normal modes can not account for this discrepancy.

In BR, the N-H band can only be assigned indirectly, via deuteration experiments. Therefore, as a final test, we computed the ν_{ND} band for the deuterated, hydrogen-bonded all-*trans* retinal (Table 7.5). In our calculations, substitution of the Schiff base's hydrogen was sufficient. An additional deuteration of the water molecule causes only minor additional shifts of about 1 cm^{-1}.

Unfortunately, the results of our analysis of the ν_{NH} frequencies are inconclusive. While the absolute frequencies for the N-H/N-D stretch vibrations are described well with respect to methods of higher accuracy, the same cannot be said for the shifts due to hydrogen bonding. Since the methods show deviating behavior with respect to each other, and experimental information is unavailable, a conclusion cannot be made at this point. A more rigorous analysis of the vibrations, especially in combination with C=C, C-H and O-H stretches might reveal additional information, that might help to estimate the true performance of DFTB.

7.3 Analysis of the N-H Stretch Vibration in Bacteriorhodopsin

In the difference FTIR spectrum of BR, the N-D stretching vibration was assigned to 2123 cm^{-1} in the dark state and 2468 cm^{-1} in the K intermediate.[293] The blueshift of 345 cm^{-1} indicates that the hydrogen-bonding interactions of the Schiff base become weaker in going from the dark state to K. Unfortunately, intensities are not available in the current implementation of DFTB/CHARMM, so that a precise assignment of the N-H vibration is impossible at the moment and has to be approximated by visual inspection of the normal modes. The frequency ranges of the N-H bond stretches are shown in Table 7.6. Due to the large isotope shift of the N-H vibration (Fig. 7.3), the N-D vibration is clearly distinct from the C-H and O-H stretches. The blueshifts of the ν_{ND} stretch vibration are obtained by both DFTB variants qualitatively. The shift obtained with DFTB2 is 373 cm^{-1}, therefore in good agreement with the experiment, while DFTB3 seems to underestimate the difference (185 cm^{-1}). A normal mode analysis performed with a different method obtains 2113 cm^{-1} for the N-D stretch, which is in good agreement with DFTB2.[294]

As mentioned above, an experimental assignment of the N-H stretch frequency is difficult and occurs through very indirect argumentation, e.g. by comparison to model compounds.[295] Therefore, the referenced N-H frequencies are best regarded as qualified estimates. In light of that, the computed N-H and N-D stretch frequencies reasonably describe the experimental situation. While applying all due caution, both the isotope shifts as well as the shifts from BR to K indicate, that with DFTB, the N-H stretch vibration is not shifted to frequencies below 3000 cm^{-1}, so that it can not account for

Table 7.6 Frequency ranges of the normal modes with contributions of the N-H bond stretch in BR and K (in cm^{-1}). The Schiff base proton and the three water molecules 401, 402 and 406 were deuterated.

vibration	structure	DFTB2	DFTB3	expt.[287,293]
N – H	BR	3143 – 3289	3012 – 3104	≈3380
N – D	BR	2193	2215,2272	2123
N – H	K	3503	3494	3310 – 3350
N – D	K	2566	2428	2468

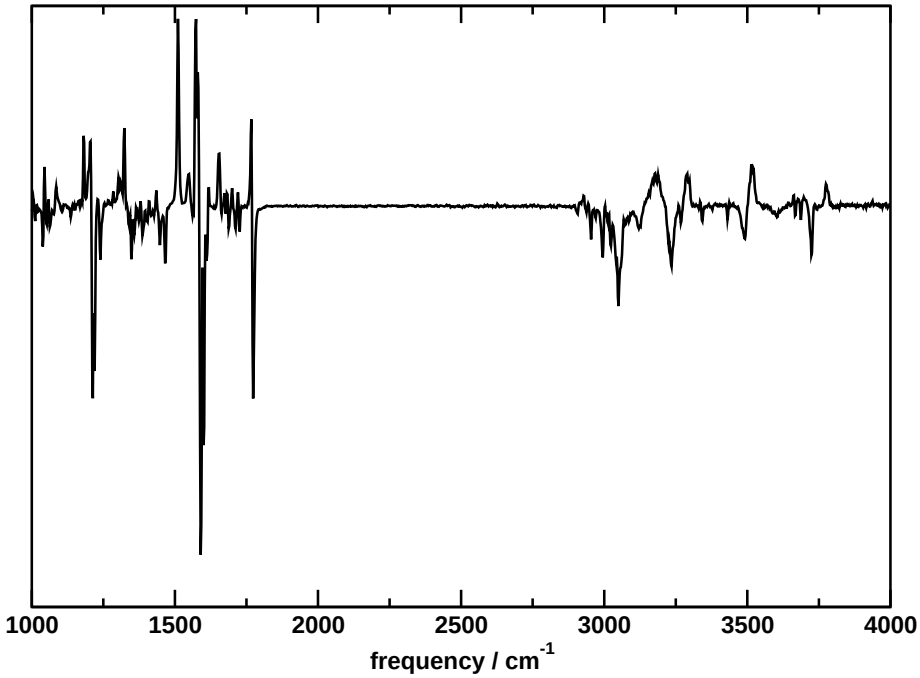


Figure 7.5 Simulated FTIR difference spectrum of K minus BR. Simulated with DFTB3/CHARMM

the occurrence of the broad and weak band in the FTIR spectrum of the dark state of BR. In the gas phase, DFTB3 underestimates the shifts due to hydrogen bonding (Fig. 7.4) with respect to DFTB2. In bacteriorhodopsin, no significant differences are observed, when replacing DFTB3 with DFTB2.

To test, whether a flexible hydrogen bond of the Schiff base with the water molecule causes a shape of the spectrum, which is similar to the weak band observed from the difference spectrum, we also simulated the FTIR spectrum via the FTTCF approach from QM/MM MD simulations of both the dark and K states. (Fig. 7.5).

In the simulated difference spectrum, no broad band could be observed. Unfortunately, an assignment of the N-H vibration to bands of the spectrum is not immediately possible. Although methods exist, that try to localize the normal modes,[296] their application is not straightforward. In the vibrational spectra of the gas phase models, the ν_{NH} band was always highest in intensity in its spectral region (3000 – 3500 cm^{-1}). For BR, DFTB3 yields a range of the ν_{NH} stretch frequencies of 3012 – 3014 cm^{-1} (Table 7.6). In the FTTCF spectrum, the band which has highest intensity has its peak at ≈3100 $cm^{-1}(-)$ which can be tentatively assigned to the N-H stretch vibration. The band is broad,

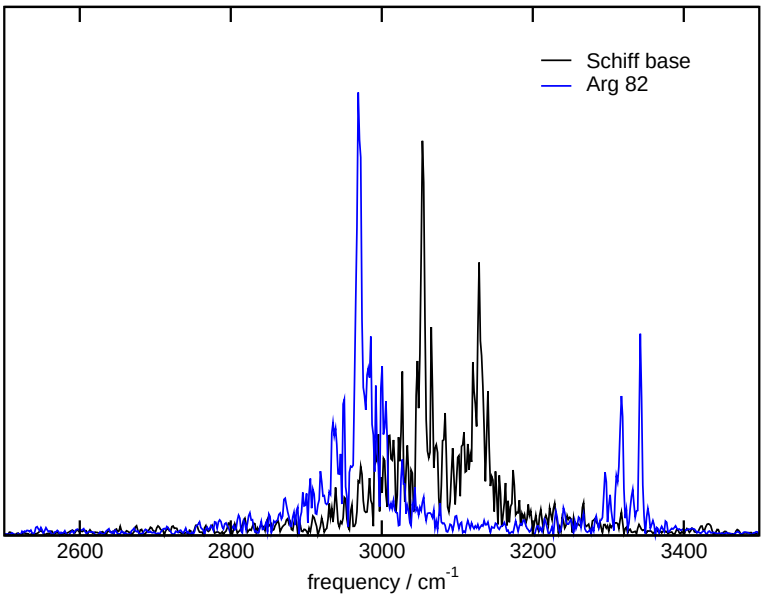


Figure 7.6 Fourier-transformed N-H bond length time series from the Schiff base and the $H_{\eta 1}$ of Arg 82.

because it overlaps with the many C-H stretch vibrations of the retinal, which occupy the frequencies 2800 – 3000 cm^{-1}. Neither increasing the QM region nor using DFTB2 as QM method caused significant changes in the spectrum.

Baer et al. reported a good agreement of their FTIR spectra, if the Arg 82 is included in the QM region.[297] However, their comparison of the N-H stretch vibration is based on the comparison of the Fourier-transformed of the N-H bond length time series (power spectrum), which is not the same as the dipole moments (cf. Fig. 3 of ref. 297). It is clear, that the contribution below 3000 cm^{-1} in the power spectrum comes from Arg 82 and not from the Schiff base (cf. Fig. 4 of ref. 297 and Fig. 7.6). Furthermore, their structural ensemble on which these calculations are based upon contains a hydrogen bond between Arg 82 and Asp 212, which might be an artifact of the chosen QM/MM method (cf. Fig. 2 of ref. 297), since it is neither observed in the X-ray structure of BR,[14] nor in our QM/MM simulations.

7.4 Preliminary Conclusion

Our simulations do not provide any hint as to the origin of the weak and broad band, since it is not observed in the simulated spectrum and while a normal mode analysis may not be suitable for this kind of weak spectroscopic feature, the redshift of the ν_{NH} band due to hydrogen-bonding indicates, that with the current simulation protocol, it is unlikely, that the hydrogen-bonded network around the Schiff base and the water molecules can be responsible for such a spectroscopic feature.

Since our calculations on the gas-phase models are inconclusive, it is unclear at this point, whether DFTB3 is underestimating the shift due to hydrogen bonding. From the consideration of shifts due to hydrogen bonding in the gas-phase models as well as the difference between the dark state of BR and K, DFTB2 behaves more systematically. The origin of the interesting discrepancy between DFTB2 and

DFTB3 certainly deserves future attention.

In the future, QM/MM normal mode analysis with DFT or simulations in the same manner as performed in ref. 297 might give a more conclusive insight on the origin of the weak and broad band in the spectrum of BR. The analysis should also be extended to other vibrations, such as O-H and C-H stretch vibrations to obtain a more complete picture.

CHAPTER 8

Conclusions and Outlook

The discovery of bacteriorhodopsin brought forward many advancements in many scientific fields, ranging from X-ray crystallography and spectroscopy to computer simulations. Since then, many other rhodopsins have been discovered and examined. What sets the channelrhodopsins apart, is their proven usefulness in neuroscientific experiments, that may one day lead to useful insights on the origin of such worrisome diseases of the brain like Parkinson's, chronic pain, autism, schizophrenia and many more. Future developments are hard to predict, but present results make believe, that some of the symptoms related to the aforementioned diseases may be treated more effectively and efficiently than currently existing protocols. Always working on the improvement of their optogenetic tools, the demand for optimized channelrhodopsin variants is steady. Targets of optimization include kinetics, sensitivity, color tuning and expression level. Especially the first three properties are subjects suitable for computational studies. They provide valuable guidelines for the manipulation of the desired properties based on mechanistic insights on the atomistic level.

In this thesis, we focussed on the connection between the structure of the protein and its spectral properties. In the first computational study on channelrhodopsins, in chapter 5 we described factors that influence the absorption spectra of the channelrhodopsins and explored the effects of single mutations on the structure and spectrum. We showed, that single mutations have only a limited effect on the absorption maximum, while the mutation of a number of amino acids in the binding pocket is indeed able to shift the spectrum by values that agree with the experimental observations. Therefore, electrostatic interactions with a polar binding pocket are able to explain the large blueshift of the absorption maximum with respect to bacteriorhodopsin. When interpreting this study, it should be kept in mind, that an X-ray structure was not available during our simulations. Recent advancements seem to support our findings that the binding pocket has the greatest effect on the absorption spectrum, while the additional number of charged residues, especially in helix 2, either do not exhibit significant effects, or their effects cancel each other. Originally, it was the hope that simulations could suggest new mutations, which shift the absorption spectrum of the channelrhodopsins to the red. However, our results show that mutation-based strategies will hardly be able to reach dimensions that are required in optogenetic experiments. Interesting successes could be reached with chimeras between two channelrhodopsins. In the future, studies of these chimeras with the help of the ChR1/ChR2 X-ray structure as template may able to explain, how the redshift compared to other channelrhodopsins is achieved.

In chapter 6 we found a better foundation for the explanation of FTIR experiments on the DC gate in ChR2. Interestingly, the assumption of (structural) homology to bacteriorhodopsin provided an explanation, that was inconsistent with newer results from computer simulations, X-ray, and their own FTIR experiments, although sequential homology was given. The reason, the homology breaks down, is that in contrast to bacteriorhodopsin, a water molecule is found in the DC gate. We showed, that QM/MM normal mode analysis is a quick and accurate way to describe the changes in the C=O stretch vibration frequency due to structural changes. The role of the water molecule in the mechanism of ChR2 is not yet clear. It is conceivable from our simulations, that it is involved *somehow* and its presence or absence may very well be responsible for the extended lifetimes of certain ChR2 mutations. Recent experiments suggest, that the DC gate is involved in the reprotonation of the Schiff base in the later part of the photocycle. It would of course be desirable to directly simulate the proton transfer in the channelrhodopsins, which requires major progesses in both experiment and simulation.

During chapter 7 we learnt, that even after 40 years of intensive studies, there are still new things to learn about bacteriorhodopsin. Encouraged by the successful application of DFTB in the QM/MM vibrational analysis of the DC gate in ChR2, we investigated possible origins of a broad and weak band around 2700 – 3100 cm^{-1} in the FTIR spectrum of bacteriorhodopsin. It is currently a qualified speculation, that the Schiff base's N-H is involved in this band. Due to hydrogen bonding with a water molecule, the frequency of the N-H stretch could get shifted near 3000 cm^{-1}. Our simulations showed two things so far: first, the N-H stretch vibration is more difficult to describe correctly, as performance of DFTB differs more from more accurate methods than it was the case for the C=O vibration. Experimental data on the N-H stretch in protonated Schiff bases in the gas phase is not available for comparison, so that it is difficult to truly judge. Our simulations, both static normal mode analysis and spectra obtained from molecular dynamics, did not show any hint as to the origin of this band. Future inquiries should include QM/MM investigations with DFT in order to rule out deficiencies of DFTB and combine the analyses with other vibrations, such as the C=C, C-H, and O-H vibrations.

In studying the active site of the channelrhodopsins, progress is hampered by limitations of current methodology. The properties of the protein, such as hydrogen-bonded networks, water density or dynamic stability are influenced by small changes, e.g. the protonation of a titrable amino acid. Therefore, interactions have to be described accurately and more important, unbiased. In bacteriorhodopsin, where the structure of the active site is known in great detail, it is known that classical force fields favor the direct interaction between positively and negatively charged residues over interactions with a water molecule. It is also known, that a better description is achieved, if explicit polarization of the amino acids is taken into account. On the other hand, conventional QM/MM implementations are too slow to achieve any relevant amount of sampling for unbiased structural ensembles of the active site. Further, the flexibility of water molecules prevents application of standard QM/MM schemes, since they require a fixed definition of QM residues a priori. Both topics are worked on actively, passed first tests, and will improve future studies of structure and spectroscopy in the channelrhodopsins.

Part IV

Appendix

APPENDIX A

Microbial Opsin Sequences

The full list of amino acid sequences used to create the phylogenetic tree shown in Figure 2.1 (p 12) are provided here.

Table A.1 Opsin sequences found in *C. reinhardtii*, their abbreviations, and reports in the literature. The accession number refers to the code of the amino acid sequence, which is deposited in the NCBI GenBank.[55]

organism	name(s)	abbreviation	accession no.
Chlamydomonas reinhardtii	chlamyopsin, chlamyopsin-1*	cop, cop1[43]	CAA88832
	chlamyopsin-2*	cop2[45]	AAG02503
	channelrhodopsin-1, chlamyopsin-3,	ChR1[49], cop3[48]	AAL08946
	Chlamydomonas sensory opsin A,	CSOA[51]	
	archaeal-type Chlamydomonas opsin 1	Acop1[56]	
	channelrhodopsin-2, chlamyopsin-4,	ChR2[50], cop4[48]	AAM15777
	Chlamydomonas sensory opsin B,	CSOB[51]	
	archaeal-type Chlamydomonas opsin 2	Acop2[56]	
	histidine kinase rhodopsin†,chlamyopsin-5	hkr1,cop5[52,54]	AAQ16277
	chlamyopsin-6	cop6[52,53]	EDO99289
	chlamyopsin-7	cop7[52]	
Volvox carterii	volvoxopsin-1	vop1 [298,299]	CAA72088
	channelrhodopsin-1	VChR1[62,219]	ABZ90901
	channelrhodopsin-2	VChR2[219]	ABZ90903
	volvoxopsin-5	vop5[299]	EFJ44204
	volvoxopsin-6	vop6[299]	EFJ41867

Pleodorina starii	channelrhodopsin-1	PsChR1[60]	AEY68812
	channelrhodopsin-2	PsChR2[60]	AEY68813
	pleopsin-5	PsCop5[60]	AEY68814
	pleopsin-6	PsCop6[60]	AEY68815
Dunaliella salina	channelrhodopsin-1	DChR1[60]	AEY68833
Mesostigma viride	channelrhodopsin-1	MChR1[61]	AEI83869
Pyramimonas gelidicola	channelrhodopsin-1	PgChR1[60]	AEY68835
Halobacterium salinarum	bacteriorhodopsin	BR[300]	CAP14056
	halorhodopsin	HR[301]	BAA07823
	sensory rhodopsin I	sRI[300]	CAP14202
	sensory rhodopsin II	sRII[300]	CAP14279
Anabaena sp.	Anabena sensory rhodopsin	ASR[302]	1XIO_A
Acetabularia acetabulum	rhodopsin	AaRh[303]	AAY82897
Chlorella vulgaris	rhodopsin	CvRh[60]	AEY68816
Guillardia theta	rhodopsin	Gt1Rh[304]	ABA08437
	rhodopsin-2	Gt2Rh[304]	ABA08438
	rhodopsin	GtR3[305]	AEY68834
Podospora Anserine	bacteriorhodopsin-like	Pop[306]	CAP62060
Leptosphaeria maculans	*L. maculans* rhodopsin	Mac[307]	AAG01180
Neurospora crassa	rhodopsin-1	nop-1[308]	EAA30185
Haloarcula argentinensis	cruxhalorhodopsin-1	CHR-1[309]	Q53461
	cruxrhodopsin-1	cR1[309]	BAA06678
Halorubrum sodomense	archaerhodopsin-3	aR-3/Arch[310]	BAA09452
Haloarcula marismortui	halorhodopsin	cHR-5[311]	AAV46572
Cryptomonas sp.	rhodopsin	CsRh[304]	ABA08439
Cyanophora paradoxa	rhodopsin	OpsCp[312]	ACV05065
Krokinobacter eikastus	proteorhodopsin	KR1[313]	BAN14807
	sodium pumping rhodopsin	KR2[313]	BAN14808

∗ cop1 and cop2 originate from the same gene and are alternative splicing products.

† cop5 was originally deposited in 2004 and was later determined to be the HKR. The name cop5 was assigned anew in 2007, under the accession number EDP06598. Its sequence is 98 % homologous to HKR.

Publications and Copyright

List of Publications

1. Wolter, T.; Welke, K.; Phatak, P.; Bondar, A.-N.; Elstner, M. "Excitation energies of a water-bridged twisted retinal structure in the bacteriorhodopsin proton pump: a theoretical investigation." *Phys. Chem. Chem. Phys.*, **2013**, 15, 12582–90.

2. Watanabe, H. C.; Welke, K.; Sindhikara, D. J.; Hegemann, P.; Elstner, M. "Towards an Understanding of Channelrhodopsin Function: Simulations Lead to Novel Insights of the Channel Mechanism." *J. Mol. Biol.*, **2013**, 425, 1795–814.

3. Welke, K.; Watanabe, H. C.; Wolter, T.; Gaus, M.; Elstner, M. "QM/MM simulations of vibrational spectra of bacteriorhodopsin and channelrhodopsin-2." *Phys. Chem. Chem. Phys.*, **2013**, 15, 6651–9.

4. Watanabe, H. C.; Welke, K.; Schneider, F.; Tsunoda, S. P.; Zhang, F.; Deisseroth, K.; Hegemann, P.; Elstner, M. "Structural Model of Channelrhodopsin." *J. Biol. Chem.* **2012**, 287, 7456–7466.

5. Welke, K.; Frähmcke, J. S.; Watanabe, H. C.; Hegemann, P.; Elstner, M. "Color Tuning in Binding Pocket Models of the Chlamydomonas-Type Channelrhodopsins." *J. Phys. Chem. B* **2011**, 115, 15119–28.

6. Moore, E. B.; De La Llave, E.; Welke, K.; Scherlis, D. A.; Molinero, V. "Freezing, melting and structure of ice in a hydrophilic nanopore." *Phys. Chem. Chem. Phys.* **2010**, 12, 4124–4134.

Copyright

Abbreviations

AO	atomic orbital
ATP	adenosine triphosphate
BLA	bond length alternation
BR	bacteriorhodopsin
CAS	complete active space
CD	circular dichroism
CI	configuration interaction
CHARMM	Chemistry at Harvard Macromolecular Mechanics
ChR	channelrhodopsin
DFT	density functional theory
DFTB	density functional tight-binding
DNA	desoxyribonucleic acid
FTIR	Fourier-transformed infrared
FTTCF	Fourier-transformed time correlation function
GGA	generalized gradient approximation
GPCR	G-protein coupled receptor
HF	Hartree-Fock
LCAO	linear combination of atomic orbitals
LDA	local density approximation
MD	molecular dynamics
MO	molecular orbital
MP2	Møller-Plesset perturbation theory, 2nd order
NMR	nuclear magnetic resonance
OM2	orthogonalization model 2
PME	particle mesh Ewald
PRG	proton release group
RAS	restricted active space
SCF	self-consistent field
SORCI	spectroscopy oriented configuration interaction
QM/MM	quantum mechanics/molecular mechanics
TIP3P	transferable interaction potential with 3 point charges

Bibliography

[1] Spudich, J. L.; Yang, C. S.; Jung, K. H.; Spudich, E. N. *Annu. Rev. Cell Dev. Biol.* **2000**, *16*, 365–92.

[2] Terakita, A. *Genome Biol.* **2005**, *6*, 213.

[3] McBee, J. K.; Palczewski, K.; Baehr, W.; Pepperberg, D. R. *Prog. Retinal Eye Res.* **2001**, *20*, 529–469.

[4] Shen, L.; Chen, C.; Zheng, H.; Jin, L. *Sci. World J.* **2013**, *2013*, 435651.

[5] Devine, E. L.; Oprian, D. D.; Theobald, D. L. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 13351–13355.

[6] Oesterhelt, D.; Stoeckenius, W. *Nature* **1971**, *233*, 149–152.

[7] Woese, C. R.; Kandlert, O.; Wheelis, M. L. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 4576–4579.

[8] Pace, N. R. *Nature* **2006**, *441*, 289.

[9] Stoeckenius, W. *J. Cell Biol.* **1967**, *34*, 365–393.

[10] Stoeckenius, W. *J. Cell Biol.* **1968**, *38*, 337–357.

[11] Blaurock, A. E.; Stoeckenius, W. *Nature* **1971**, *233*, 152–155.

[12] Oesterhelt, D. *Proc. Natl. Acad. Sci. U.S.A.* **1973**, *70*, 2853–2857.

[13] Henderson, R.; Unwin, P. N. T. *Nature* **1975**, *257*, 28–32.

[14] Luecke, H.; Schobert, B.; Richter, H.-T.; Cartailler, J.-P.; Lanyi, J. K. *J. Mol. Biol.* **1999**, *291*, 899–911.

[15] Lanyi, J. K. *Annu. Rev. Physiol.* **2004**, *66*, 665–88.

[16] Mathies, R. A.; Lin, S. W.; Ames, J. B.; Pollard, W. T. *Annu. Rev. Biophys. Biophys. Chem.* **1991**, *20*, 491–518.

[17] Phatak, P. V. Investigation of Proton Transfer Pathways in Bacteriorhodopsin with Multi-Length-Scale Simulations. Ph.D. thesis, TU Braunschweig, 2009.

[18] Song, L.; El-Sayed, M. A. *J. Am. Chem. Soc.* **1998**, *120*, 8889–8890.

[19] Brack, T.; Atkinson, G. *J. Mol. Struct.* **1989**, *214*, 289–303.

[20] Schobert, B.; Cupp-Vickery, J.; Hornak, V.; Smith, S. O.; Lanyi, J. K. *J. Mol. Biol.* **2002**, *321*, 715–726.

[21] Lozier, R. H.; Bogomolni, R. A.; Stoeckenius, W. *Biophys. J.* **1975**, *15*, 955–62.

[22] Nagle, J. F.; Parodi, L. A.; Lozier, R. H. *Biophys. J.* **1982**, *38*, 161–74.

[23] Stoeckenius, W.; Lozier, R. H. *J. Supramol. Struct.* **1974**, *2*, 769–74.

[24] Ludmann, K.; Gergely, C.; Váró, G. *Biophys. J.* **1998**, *75*, 3110–9.

[25] Aton, B.; Doukas, A. G.; Callender, R. H.; Becher, B.; Ebrey, T. G. *Biochemistry* **1977**, *16*, 2995–2999.

[26] Ames, J. B.; Fodor, S. P. A.; Gebhard, R.; Raap, J.; Van den Berg, E. M. M.; Lugtenburg, J.; Mathies, R. A. *Biochemistry* **1989**, *28*, 3681–3687.

[27] Gerwert, K.; Souvignier, G.; Hess, B. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 9774–9778.

[28] Bondar, A.-N.; Elstner, M.; Suhai, S.; Smith, J. C.; Fischer, S. *Structure* **2004**, *12*, 1281–8.

[29] Brown, L. S.; Sasaki, J.; Kandori, H.; Maeda, A.; Needleman, R.; Lanyi, J. K. *J. Biol. Chem.* **1995**, *270*, 27122–27126.

[30] Garczarek, F.; Gerwert, K. *Nature* **2006**, *439*, 109–112.

[31] Phatak, P.; Ghosh, N.; Yu, H.; Cui, Q.; Elstner, M. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 19672–19677.

[32] Miller, A.; Oesterhelt, D. *Biochim. Biophys. Acta, Bioenerg.* **1990**, *1020*, 57–64.

[33] Cao, Y.; Váró, G.; Klinger, A. L.; Czajkowsky, D. M.; Braiman, M. S.; Needleman, R.; Lanyi, J. K. *Biochemistry* **1993**, *32*, 1981–1990.

[34] Schobert, B.; Brown, L. S.; Lanyi, J. K. *J. Mol. Biol.* **2003**, *330*, 553–570.

[35] Smith, S. O.; Pardoen, J. A.; Mulder, P. P. J.; Curry, B.; Lugtenburg, J.; Mathies, R. *Biochemistry* **1983**, *22*, 6141–6148.

[36] Boyden, E. S.; Zhang, F.; Bamberg, E.; Nagel, G.; Deisseroth, K. *Nat. Neurosci.* **2005**, *8*, 1263–8.

[37] Watanabe, H. C.; Welke, K.; Schneider, F.; Tsunoda, S. P.; Zhang, F.; Deisseroth, K.; Hegemann, P.; Elstner, M. *J. Biol. Chem.* **2012**, *287*, 7456–7466.

[38] Kato, H. E.; Zhang, F.; Yizhar, O.; Ramakrishnan, C.; Nishizawa, T.; Hirata, K.; Ito, J.; Aita, Y.; Tsukazaki, T.; Hayashi, S.; Hegemann, P.; Maturana, A. D.; Ishitani, R.; Deisseroth, K.; Nureki, O. *Nature* **2012**, *482*, 369–374.

[39] Hegemann, P.; Nagel, G. *EMBO Mol. Med.* **2013**, *5*, 173–6.

[40] Dangeard, P. A. *Annales des Sciences Naturelles, Botanique* **1888**, *7*, 105–175.

[41] Foster, K. W.; Smyth, R. D. *Microbiol. Rev.* **1980**, *44*, 572–630.

[42] Ozawa, S.-I.; Nield, J.; Terao, A.; Stauber, E. J.; Hippler, M.; Koike, H.; Rochaix, J.-D.; Takahashi, Y. *Plant Cell* **2009**, *21*, 2424–42.

[43] Deininger, W.; Kröger, P.; Hegemann, U.; Lottspeichl, F.; Hegemann, P. *EMBO J.* **1995**, *14*, 5849–5858.

[44] Deininger, W.; Fuhrmann, M.; Hegemann, P. *Trends Genet.* **2000**, *16*, 158–159.

[45] Fuhrmann, M.; Deininger, W.; Kateriya, S.; Hegemann, P. In *Photoreceptors and Light Signalling*; Batschauer, A., Ed.; Royal Society of Chemistry: Cambridge, UK, 2003; Vol. 3.

[46] Fuhrmann, M.; Stahlberg, A.; Govorunova, E. G.; Rank, S.; Hegemann, P. *J. Cell Sci.* **2001**, *114*, 3857–3863.

[47] Asamizu, E.; Miura, K.; Kucho, K.; Inoue, Y.; Fukuzawa, H.; Ohyama, K.; Nakamura, Y.; Tabata, S. *DNA Res.* **2000**, *7*, 305–307.

[48] Hegemann, P.; Fuhrmann, M.; Kateriya, S. *J. Phycol.* **2001**, *676*, 668–676.

[49] Nagel, G.; Ollig, D.; Fuhrmann, M.; Kateriya, S.; Musti, A. M.; Bamberg, E.; Hegemann, P. *Science* **2002**, *296*, 2395–8.

[50] Nagel, G.; Szellas, T.; Huhn, W.; Kateriya, S.; Adeishvili, N.; Berthold, P.; Ollig, D.; Hegemann, P.; Bamberg, E. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 13940–5.

[51] Sineshchekov, O. A.; Jung, K.-H.; Spudich, J. L. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 8689–94.

[52] Kateriya, S.; Nagel, G.; Bamberg, E.; Hegemann, P. *News Physiol. Sci.* **2004**, *19*, 133–137.

[53] Merchant, S. S. et al. *Science* **2007**, *318*, 245–50.

[54] Luck, M.; Mathes, T.; Bruun, S.; Fudim, R.; Hagedorn, R.; Tran Nguyen, T. M.; Kateriya, S.; Kennis, J. T. M.; Hildebrandt, P.; Hegemann, P. *J. Biol. Chem.* **2012**, *287*, 40083–90.

[55] Benson, D. A.; Cavanaugh, M.; Clark, K.; Karsch-Mizrachi, I.; Lipman, D. J.; Ostell, J.; Sayers, E. W. *Nucleic Acids Res.* **2013**, *41*, D36–42.

[56] Suzuki, T.; Yamasaki, K.; Fujita, S.; Oda, K.; Iseki, M.; Yoshida, K.; Watanabe, M.; Daiyasu, H.; Toh, H.; Asamizu, E.; Tabata, S.; Miura, K.; Fukuzawa, H.; Nakamura, S.; Takahashi, T. *Biochem. Biophys. Res. Commun.* **2003**, *301*, 711–717.

[57] Larkin, M. A.; Blackshields, G.; Brown, N. P.; Chenna, R.; McGettigan, P. A.; McWilliam, H.; Valentin, F.; Wallace, I. M.; Wilm, A.; Lopez, R.; Thompson, J. D.; Gibson, T. J.; Higgins, D. G. *Bioinf.* **2007**, *23*, 2947–8.

[58] Tamura, K.; Peterson, D.; Peterson, N.; Stecher, G.; Nei, M.; Kumar, S. *Mol. Biol. Evol.* **2011**, *28*, 2731–9.

[59] Hou, S.-Y.; Govorunova, E. G.; Ntefidou, M.; Lane, C. E.; Spudich, E. N.; Sineshchekov, O. A.; Spudich, J. L. *Photochem. Photobiol.* **2012**, *88*, 119–28.

[60] Zhang, F.; Vierock, J.; Yizhar, O.; Fenno, L. E.; Tsunoda, S. P.; Kianianmomeni, A.; Prigge, M.; Berndt, A.; Cushman, J.; Polle, J.; Magnuson, J.; Hegemann, P.; Deisseroth, K. *Cell* **2011**, *147*, 1446–57.

[61] Govorunova, E. G.; Spudich, E. N.; Lane, C. E.; Sineshchekov, O. A.; Spudich, J. L. *mBio* **2011**, *2*, e00115–11.

[62] Zhang, F.; Prigge, M.; Beyrière, F.; Tsunoda, S. P.; Mattis, J.; Yizhar, O.; Hegemann, P.; Deisseroth, K. *Nat. Neurosci.* **2008**, *11*, 631–3.

[63] Müller, M.; Bamann, C.; Bamberg, E.; Kühlbrandt, W. *J. Mol. Biol.* **2011**, *414*, 86–95.

[64] Watanabe, H. C.; Welke, K.; Sindhikara, D. J.; Hegemann, P.; Elstner, M. *J. Mol. Biol.* **2013**, *425*, 1795–814.

[65] Lomize, M. a.; Lomize, A. L.; Pogozheva, I. D.; Mosberg, H. I. *Bioinformatics* **2006**, *22*, 623–5.

[66] Berthold, P.; Tsunoda, S. P.; Ernst, O. P.; Mages, W.; Gradmann, D.; Hegemann, P. *Plant Cell* **2008**, *20*, 1665–77.

[67] Tsunoda, S. P.; Hegemann, P. *Photochem. Photobiol.* **2009**, *85*, 564–9.

[68] Bamann, C.; Kirsch, T.; Nagel, G.; Bamberg, E. *J. Mol. Biol.* **2008**, *375*, 686–694.

[69] Lin, J. Y.; Lin, M. Z.; Steinbach, P.; Tsien, R. Y. *Biophys. J.* **2009**, *96*, 1803–1814.

[70] Feldbauer, K.; Zimmermann, D.; Pintschovius, V.; Spitz, J.; Bamann, C.; Bamberg, E. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 12317–22.

[71] Lin, J. Y. *Exp. Physiol.* **2011**, *96*, 19–25.

[72] Nack, M.; Radu, I.; Schultz, B.-J.; Resler, T.; Schlesinger, R.; Bondar, A.-N.; del Val, C.; Abbruzzetti, S.; Viappiani, C.; Bamann, C.; Bamberg, E.; Heberle, J. *FEBS letters* **2012**, *586*, 1344–8.

[73] Ishizuka, T.; Kakuda, M.; Araki, R.; Yawo, H. *Neurosci. Res.* **2006**, *54*, 85–94.

[74] Berndt, A.; Yizhar, O.; Gunaydin, L. A.; Hegemann, P.; Deisseroth, K. *Nature Neuroscience* **2009**, *12*, 229–234.

[75] Bamann, C.; Gueta, R.; Kleinlogel, S.; Nagel, G.; Bamberg, E. *Biochemistry* **2010**, *49*, 267–78.

[76] Hegemann, P.; Ehlenbeck, S.; Gradmann, D. *Biophys. J.* **2005**, *89*, 3911–3918.
[77] Ritter, E.; Stehfest, K.; Berndt, A.; Hegemann, P.; Bartl, F. J. *J. Biol. Chem.* **2008**, *283*, 35033–41.
[78] Lórenz-Fonfría, V. A.; Resler, T.; Krause, N.; Nack, M.; Gossing, M.; Fischer von Mollard, G.; Bamann, C.; Bamberg, E.; Schlesinger, R.; Heberle, J. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, E1273–81.
[79] Berndt, A.; Prigge, M.; Gradmann, D.; Hegemann, P. *Biophys. J.* **2010**, *98*, 753–761.
[80] Ritter, E.; Piwowarski, P.; Hegemann, P.; Bartl, F. J. *J. Biol. Chem.* **2013**, *288*, 10451–8.
[81] Zemelman, B. V.; Lee, G. A.; Ng, M.; Miesenböck, G. *Neuron* **2002**, *33*, 15–22.
[82] Lima, S. Q.; Miesenböck, G. *Cell* **2005**, *121*, 141–152.
[83] Li, X.; Gutierrez, D. V.; Hanson, M. G.; Han, J.; Mark, M. D.; Chiel, H.; Hegemann, P.; Landmesser, L. T.; Herlitze, S. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 17816–21.
[84] Deisseroth, K.; Feng, G.; Majewska, A. K.; Miesenböck, G.; Ting, A.; Schnitzer, M. J. *J. Neurosci.* **2006**, *26*, 10380–6.
[85] Fenno, L.; Yizhar, O.; Deisseroth, K. *Annu. Rev. Neurosci.* **2011**, *34*, 389–412.
[86] Pastrana, E. *Nat. Methods* **2011**, *8*, 24–25.
[87] Editorial, *Nat. Methods* **2011**, *8*, 1–1.
[88] Deisseroth, K. *Nat. Methods* **2011**, *8*, 26–9.
[89] The News Staff, *Science* **2010**, *330*, 1612–3.
[90] Zhang, F.; Gradinaru, V.; Adamantidis, A. R.; Durand, R.; Airan, R. D.; de Lecea, L.; Deisseroth, K. *Nat. Protoc.* **2010**, *5*, 439–56.
[91] Nagel, G.; Brauner, M.; Liewald, J. F.; Adeishvili, N.; Bamberg, E.; Gottschalk, A.
[92] Zhang, F.; Wang, L.-P.; Brauner, M.; Liewald, J. F.; Kay, K.; Watzke, N.; Wood, P. G.; Bamberg, E.; Nagel, G.; Gottschalk, A.; Deisseroth, K. *Nature* **2007**, *446*, 633–9.
[93] Rein, M. L.; Deussing, J. M. *Mol. Genet. Genom.* **2012**, *287*, 95–109.
[94] Yizhar, O. et al. *Nature* **2011**, *477*, 171–8.
[95] Bi, A.; Cui, J.; Ma, Y.-P.; Olshevskaya, E.; Pu, M.; Dizhoor, A. M.; Pan, Z.-H. *Neuron* **2006**, *50*, 23–33.
[96] Nirenberg, S.; Pandarinath, C. *Proc. Natl. Acad. Sci. U.S.A.* **2012**,
[97] Kringelbach, M. L.; Jenkinson, N.; Owen, S. L. F.; Aziz, T. Z. *Nat. Rev. Neurosci.* **2007**, *8*, 623–35.
[98] LaLumiere, R. T. *Brain Stimul.* **2011**, *4*, 1–6.
[99] Tye, K. M.; Deisseroth, K. *Nat. Rev. Neurosci.* **2012**, *13*, 251–66.
[100] Lin, J. Y.; Knutsen, P. M.; Muller, A.; Kleinfeld, D.; Tsien, R. Y. *Nat. Neurosci.* **2013**, *16*, 1499–508.
[101] Eisenhauer, K.; Kuhne, J.; Ritter, E.; Berndt, A.; Wolf, S.; Freier, E.; Bartl, F. J.; Hegemann, P.; Gerwert, K. *J. Biol. Chem.* **2012**, *287*, 6904–6911.
[102] Jensen, F. *Introduction to Computational Chemistry*, 2nd ed.; John Wiley & Sons: West Sussex, England, 2007.
[103] Koch, W.; Holthausen, M. C. *A Chemist's Guide to Density Functional Theory*, 2nd ed.; Wiley-VCH: Weinheim, Germany, 2007.
[104] Leach, A. R. *Molecular Modelling: Principles and Applications*; Pearson Publication, 2001.
[105] Szabo, A.; Ostlund, N. S. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*; Dover Publications, 1996.

[106] Mulliken, R. S. *J. Chem. Phys.* **1955**, *23*, 1833.
[107] Gordon, R. G. *J. Chem. Phys.* **1965**, *43*, 1307.
[108] Kubo, R. *J. Phys. Soc. Jpn.* **1957**, *12*, 570–586.
[109] McQuarrie, D. A. *Statistical Mechanics*; University Science Books, 2000.
[110] Ramírez, R.; López-Ciudad, T.; Kumar P, P.; Marx, D. *J. Chem. Phys.* **2004**, *121*, 3973–83.
[111] Hohenberg, P. *Phys. Rev.* **1964**, *136*, B864–B871.
[112] Wang, Y. A.; Carter, E. A. In *Theoretical Methods in Condensed Phase Chemistry*; Schwartz, S. D., Ed.; Kluwer: Dordrecht, 2000; Chapter 5, pp 117–183.
[113] Parr, R. G.; Yang, W. *Density-Functional Theory of Atoms and Molecules*; Oxford University Press, USA, 1994.
[114] Kohn, W.; Sham, L. J. *Phys. Rev.* **1965**, *140*, A1133–A1138.
[115] Eichkorn, K.; Treutler, O.; Öhm, H.; Häser, M.; Ahlrichs, R. *Chem. Phys. Lett.* **1995**, *240*, 283–290.
[116] Skylaris, C.-K.; Gagliardi, L.; Handy, N.; Ioannou, A.; Spencer, S.; Willetts, A. *J. Mol. Struct.: THEOCHEM* **2000**, *501*, 229–239.
[117] Gaus, M. Extension and Parametrization of an Approximate Density Functional Method for Organic and Biomolecules. PhD thesis, Karlsruhe Institute of Technology, 2011.
[118] Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.
[119] Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098–3100.
[120] Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785–789.
[121] Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648–5652.
[122] Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. *Phys. Rev. B* **1998**, *58*, 7260–7268.
[123] Gaus, M.; Cui, Q.; Elstner, M. *WIREs Comput. Mol. Sci.* **2014**, *4*, 49–61.
[124] Porezag, D.; Frauenheim, T.; Köhler, T.; Seifert, G.; Kaschner, R. *Phys. Rev. B* **1995**, *51*, 12947–12957.
[125] Seifert, G. *J. Phys. Chem. A* **2007**, *111*, 5609–5613.
[126] Gaus, M.; Cui, Q.; Elstner, M. *J. Chem. Theory Comput.* **2011**, *7*, 931–948.
[127] Kubař, T.; Bodrog, Z.; Gaus, M.; Köhler, C.; Aradi, B.; Frauenheim, T.; Elstner, M. *J. Chem. Theory Comput.* **2013**, *9*, 2939–2949.
[128] Parr, R. G.; Pearson, R. G. *J. Am. Chem. Soc.* **1983**, *105*, 7512–7516.
[129] Yang,; Yu, H.; York, D.; Cui, Q.; Elstner, M. *J. Phys. Chem. A* **2007**, *111*, 10861–10873.
[130] Langhoff, S. R.; Davidson, E. R. *Int. J. Quant. Chem.* **1974**, *8*, 61–72.
[131] Andersson, K.; Malmqvist, P.-A.; Roos, B. O. *J. Chem. Phys.* **1992**, *96*, 1218.
[132] Pulay, P. *Int. J. Quant. Chem.* **2011**, *111*, 3273–3279.
[133] Neese, F. *J. Chem. Phys.* **2003**, *119*, 9428.
[134] Mogensen, B. J.; Rettrup, S. *Int. J. Quant. Chem.* **1992**, *44*, 1045–1056.
[135] Miralles, J.; Castell, O.; Caballol, R.; Malrieu, J.-P. *Chem. Phys.* **1993**, *172*, 33–43.
[136] Weber, W.; Thiel, W. *Theor. Chem. Acc.* **2000**, *103*, 495–506.
[137] Koslowski, A.; Beck, M. E.; Thiel, W. *J. Comput. Chem.* **2003**, *24*, 714–26.

[138] MacKerell Jr., A. D. et al. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
[139] MacKerell Jr., A. D.; Feig, M.; Brooks III., C. L. *J. Am. Chem. Soc.* **2004**, *126*, 698–699.
[140] Feller, S. E.; MacKerell Jr., A. D. *J. Phys. Chem. B* **2000**, *104*, 7510–7515.
[141] Feller, S. E.; Gawrisch, K.; MacKerell Jr., A. D. *J. Am. Chem. Soc.* **2002**, *124*, 318–326.
[142] Schlenkrich, M.; Brickmann, J.; Mackerell Jr., A. D.; Karplus, M. In *Biological Membranes: A Molecular Perspective from Computation and Experiment*; Merz, K. M., Roux, B., Eds.; Birkham: Boston, 1996; Chapter Empirical.
[143] Brooks, B. R. et al. *J. Comput. Chem.* **2009**, *30*, 1545–614.
[144] Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089.
[145] Hockney, R.; Goel, S.; Eastwood, J. *J. Comput. Phys.* **1974**, *14*, 148–158.
[146] Swope, W. C. *J. Chem. Phys.* **1982**, *76*, 637.
[147] Nosé, S. *Mol. Phys.* **1984**, *52*, 255–268.
[148] Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695–1697.
[149] Parrinello, M.; Rahman, A. *J. Appl. Phys.* **1981**, *52*, 7182–7190.
[150] Nosé, S.; Klein, M. *Mol. Phys.* **1983**, *50*, 1055–1076.
[151] Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227–249.
[152] Lin, H.; Truhlar, D. G. *Theor. Chem. Acc.* **2007**, *117*, 185–199.
[153] Martin, H.; Thiel, W. In *Atomistic Approaches in Modern Biology*; Reiher, M., Ed.; Springer: Heidelberg, 2007; pp 173–290.
[154] Senn, H. M.; Thiel, W. *Angew. Chem. Intl. Ed.* **2009**, *48*, 1198–229.
[155] Cui, Q.; Elstner, M.; Kaxiras, E.; Frauenheim, T.; Karplus, M. *J. Phys. Chem. B* **2001**, *105*, 569–585.
[156] König, P. H.; Hoffmann, M.; Frauenheim, T.; Cui, Q. *J. Phys. Chem. B* **2005**, *109*, 9082–95.
[157] Hou, G.; Zhu, X.; Elstner, M.; Cui, Q. *J. Chem. Theory Comput.* **2012**, *8*, 4293–4304.
[158] Hoffmann, M.; Wanko, M.; Strodel, P.; König, P. H.; Frauenheim, T.; Schulten, K.; Thiel, W.; Tajkhorshid, E.; Elstner, M. *J. Am. Chem. Soc.* **2006**, *128*, 10808–10818.
[159] Wanko, M.; Hoffmann, M.; Frauenheim, T.; Elstner, M. *J. Phys. Chem. B* **2008**, *112*, 11462–7.
[160] Welke, K.; Frähmcke, J. S.; Watanabe, H. C.; Hegemann, P.; Elstner, M. *J. Phys. Chem. B* **2011**, *115*, 15119–28.
[161] Wanko, M.; Hoffmann, M.; Frähmcke, J. S.; Frauenheim, T.; Elstner, M. *J. Phys. Chem. B* **2008**, *112*, 11468–78.
[162] Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269–10280.
[163] Ren, P.; Ponder, J. W. *J. Phys. Chem. B* **2003**, *107*, 5933–5947.
[164] Thole, B. *Chem. Phys.* **1981**, *59*, 341–350.
[165] Luzhkov, V.; Warshel, A. *J. Am. Chem. Soc.* **1991**, *113*, 4491–4499.
[166] Thompson, M. A.; Schenter, G. K. *J. Phys. Chem.* **1995**, *99*, 6374–6386.
[167] Frähmcke, J. S.; Wanko, M.; Elstner, M. *J. Phys. Chem. B* **2012**, *116*, 3313–3321.
[168] Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. *Proteins* **2010**, *78*, 1950–8.
[169] Yu, H.; van Gunsteren, W. F. *Comput. Phys. Commun.* **2005**, *172*, 69–85.

[170] Duan, L. L.; Mei, Y.; Zhang, Q. G.; Zhang, J. Z. H. *J. Chem. Phys.* **2009**, *130*, 115102.
[171] Tong, Y.; Ji, C. G.; Mei, Y.; Zhang, J. Z. H. *J. Am. Chem. Soc.* **2009**, *131*, 8636–8641.
[172] Ji, C. G.; Zhang, J. Z. H. *J. Phys. Chem. B* **2009**, *113*, 13898–13900.
[173] Wei, C.; Tung, D.; Yip, Y. M.; Mei, Y.; Zhang, D. *J. Chem. Phys.* **2011**, *134*, 171101.
[174] Sun, T.; Wei, C.; Neo, N. W. C.; Zhang, D. *Theor. Chem. Acc.* **2013**, *132*, 1354.
[175] Duan, L. L.; Mei, Y.; Zhang, D.; Zhang, Q. G.; Zhang, J. Z. H. *J. Am. Chem. Soc.* **2010**, *132*, 11159–64.
[176] Ji, C. G.; Zhang, J. Z. H. *J. Phys. Chem. B* **2009**, *113*, 16059–16064.
[177] Wroblewska, L.; Skolnick, J. *J. Comput. Chem.* **2007**, *28*, 2059–66.
[178] Paschek, D.; Hempel, S.; García, A. E. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 17754–9.
[179] Raval, A.; Piana, S.; Eastwood, M. P.; Dror, R. O.; Shaw, D. E. *Proteins: Struct., Funct., Bioinf.* **2012**, *80*, 2071–2079.
[180] Gao, Y.; Guo, M.; Mei, Y.; Zhang, J. Z. H. *Molecular Physics* **2012**, *110*, 595–604.
[181] Tong, Y.; Mei, Y.; Li, Y. L.; Ji, C. G.; Zhang, J. Z. H. *J. Am. Chem. Soc.* **2010**, *132*, 5137–42.
[182] Mei, Y.; Li, Y. L.; Zeng, J.; Zhang, J. Z. H. *J. Comput. Chem.* **2012**, *33*, 1374–82.
[183] Kuhn, B.; Kollman, P. A. *J. Am. Chem. Soc.* **2000**, *122*, 3909–3916.
[184] Söderhjelm, P.; Ryde, U. *J. Comput. Chem.* **2009**, *30*, 750–760.
[185] Chang, L.; Ishikawa, T.; Kuwata, K.; Takada, S. *J. Comput. Chem.* **2013**, *34*, 1251–7.
[186] Illingworth, C. J. R.; Morris, G. M.; Parkes, K. E. B.; Snell, C. R.; Reynolds, C. A. *J. Phys. Chem. A* **2008**, *112*, 12157–12163.
[187] Liu, J.; He, X.; Zhang, J. Z. H. *J. Chem. Inf. Model.* **2013**,
[188] Yao, X. X.; Ji, C. G.; Xie, D. Q.; Zhang, J. Z. H. *J. Comput. Chem.* **2013**, *34*, 1136–42.
[189] Ji, C.; Mei, Y.; Zhang, J. Z. H. *Biophys. J.* **2008**, *95*, 1080–8.
[190] Simonson, T.; Carlsson, J.; Case, D. A. *J. Am. Chem. Soc.* **2004**, *126*, 4167–4180.
[191] Jämbeck, J. P. M.; Mocci, F.; Lyubartsev, A. P.; Laaksonen, A. *J. Comput. Chem.* **2013**, *34*, 187–97.
[192] Babitzki, G.; Denschlag, R.; Tavan, P. *J. Phys. Chem. B* **2009**, *113*, 10483–95.
[193] Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.
[194] Babitzki, G.; Mathias, G.; Tavan, P. *J. Phys. Chem. B* **2009**, *113*, 10496–10508.
[195] Rieff, B.; Bauer, S.; Mathias, G.; Tavan, P. *J. Phys. Chem. B* **2011**, *115*, 11239–53.
[196] Dal Peraro, M.; Spiegel, K.; Lamoureux, G.; De Vivo, M.; DeGrado, W. F.; Klein, M. L. *J. Struct. Biol.* **2007**, *157*, 444–53.
[197] Lu, Y.; Mei, Y.; Zhang, J. Z. H.; Zhang, D. *J. Chem. Phys.* **2010**, *132*, 131101.
[198] Zhu, T.; Xiao, X.; Ji, C.; Zhang, J. Z. H. *J. Chem. Theory Comput.* **2013**, *9*, 1788–1798.
[199] Beauchamp, K. A.; Lin, Y.-S.; Das, R.; Pande, V. S. *J. Chem. Theory Comput.* **2012**, *8*, 1409–1414.
[200] Orozco, M.; Luque, F. J. *Chem. Rev.* **2000**, *100*, 4187–4226.
[201] Patel, S.; Brooks, C. L. *Mol. Sim.* **2006**, *32*, 231–249.
[202] Warshel, A.; Kato, M.; Pisliakov, A. V. *J. Chem. Theory Comput.* **2007**, *3*, 2034–2045.
[203] Xie, W.; Gao, J. *J. Chem. Theory Comput.* **2007**, *3*, 1890–1900.
[204] Giese, T. J.; Chen, H.; Dissanayake, T.; Giambaşu, G. M.; Heldenbrand, H.; Huang, M.; Kuech-

ler, E. R.; Lee, T.-S.; Panteva, M. T.; Radak, B. K.; York, D. M. *J. Chem. Theory Comput.* **2013**, *9*, 1417–1427.

[205] Nakanishi, K.; Balogh-Nair, V.; Arnaboldi, M.; Tsujimoto, K.; Honig, B. *J. Am. Chem. Soc.* **1980**, *102*, 7945–7947.

[206] Kochendoerfer, G. G.; Lin, S. W.; Sakmar, T. P.; Mathies, R. A. *Trends Biochem. Sci.* **1999**, *24*, 300–305.

[207] Wanko, M.; Hoffmann, M.; Strodel, P.; Koslowski, A.; Thiel, W.; Neese, F.; Frauenheim, T.; Elstner, M. *J. Phys. Chem. B* **2005**, *109*, 3606–15.

[208] Valsson, O.; Angeli, C.; Filippi, C. *Phys. Chem. Chem. Phys.* **2012**, *14*, 11015–20.

[209] Valsson, O.; Filippi, C. *J. Phys. Chem. Lett.* **2012**, *3*, 908–912.

[210] Walczak, E.; Szefczyk, B.; Andruniów, T. *J. Chem. Theory Comput.* **2013**, 4915–4927.

[211] Andersen, L. H.; Nielsen, I. B.; Kristensen, M. B.; El Ghazaly, M. O. A.; Haacke, S.; Nielsen, M. B. n.; Petersen, M. A. *J. Am. Chem. Soc.* **2005**, *127*, 12347–50.

[212] Rajput, J.; Rahbek, D. B.; Andersen, L. H.; Hirshfeld, A.; Sheves, M.; Altoè, P.; Orlandi, G.; Garavelli, M. *Angew. Chem. Intl. Ed.* **2010**, *49*, 1790–3.

[213] Sneskov, K.; Olsen, J. M. H.; Schwabe, T.; Hättig, C.; Christiansen, O.; Kongsted, J. *Phys. Chem. Chem. Phys.* **2013**, *15*, 7567–76.

[214] Altoè, P.; Cembran, A.; Olivucci, M.; Garavelli, M. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 20172–7.

[215] Frähmcke, J. S.; Wanko, M.; Phatak, P.; Mroginski, M. A.; Elstner, M. *J. Phys. Chem. B* **2010**, *114*, 11338–11352.

[216] Wolter, T.; Welke, K.; Phatak, P.; Bondar, A.-N.; Elstner, M. *Phys. Chem. Chem. Phys.* **2013**, *15*, 12582–90.

[217] Nack, M.; Radu, I.; Bamann, C.; Bamberg, E.; Heberle, J. *FEBS Letters* **2009**, *583*, 3676–3680.

[218] Stehfest, K.; Hegemann, P. *ChemPhysChem* **2010**, *11*, 1120–6.

[219] Kianianmomeni, A.; Stehfest, K.; Nematollahi, G.; Hegemann, P.; Hallmann, A. *Plant Physiol.* **2009**, *151*, 347–66.

[220] Birge, R. R.; Zhang, C.-F. *J. Chem. Phys.* **1990**, *92*, 7178.

[221] Hayashi, S.; Tajkhorshid, E.; Pebay-Peyroula, E.; Royant, A.; Landau, E. M.; Navarro, J.; Schulten, K. *J. Phys. Chem. B* **2001**, *105*, 10124–10131.

[222] Sasaki, J.; Lanyi, J. K.; Needleman, R.; Yoshizawa, T.; Maeda, A. *Biochemistry* **1994**, *33*, 3178–3184.

[223] Dinner, A. R.; Lopez, X.; Karplus, M. *Theor. Chem. Acc.* **2003**, *109*, 118–124.

[224] Elstner, M.; Cui, Q. In *Solvation Effects. Methods and Applications*; Canuto, S., Ed.; Springer Series: Challenges and Advances in Computational Chemistry and Physics; Springer, 2008; Chapter Combined Q.

[225] Cui, Q.; Elstner, M. In *Multiscale Quantum Models for Biocatalysis: Modern Techniques and Applications*; Lee, T., York, D., Eds.; Springer, 2008; Chapter Multiscale.

[226] Zhou, H.; Tajkhorshid, E.; Frauenheim, T.; Suhai, S.; Elstner, M. *Chem. Phys.* **2002**, *277*, 91–103.

[227] Thiel, W. *MNDO99*; Version 6.1; Max-Planck-Institut für Kohlenforschung: Mülheim an der Ruhr, Germany, 2006.

[228] Neese, F. *WIREs Comput. Mol. Sci.* **2012**, *2*, 73–78.

[229] Schäfer, A.; Horn, H.; Ahlrichs, R. *J. Chem. Phys.* **1992**, *97*, 2571.

[230] Phatak, P.; Frähmcke, J. S.; Wanko, M.; Hoffmann, M.; Strodel, P.; Smith, J. C.; Suhai, S.; Bondar, A.-N.; Elstner, M. *J. Am. Chem. Soc.* **2009**, *131*, 7064–7078.

[231] Guex, N.; Peitsch, M. C. *Electrophoresis* **1997**, *18*, 2714–23.

[232] Arnold, K.; Bordoli, L.; Kopp, J.; Schwede, T. *Bioinformatics* **2006**, *22*, 195–201.

[233] Bordoli, L.; Kiefer, F.; Arnold, K.; Benkert, P.; Battey, J.; Schwede, T. *Nat. Protoc.* **2009**, *4*, 1–13.

[234] Nack, M.; Radu, I.; Gossing, M.; Bamann, C.; Bamberg, E.; von Mollard, G. F.; Heberle, J. *Photochem. Photobiol. Sci.* **2010**, *9*, 194–8.

[235] Christiansen, O.; Koch, H.; Jø rgensen, P. *Chem. Phys. Lett.* **1995**, *243*, 409–418.

[236] Hättig, C.; Weigend, F. *J. Chem. Phys.* **2000**, *113*, 5154.

[237] Kamiya, M.; Kato, H. E.; Ishitani, R.; Nureki, O.; Hayashi, S. *Chem. Phys. Lett.* **2013**, *556*, 266–271.

[238] Schwabe, T.; Sneskov, K.; Haugaard Olsen, J. M.; Kongsted, J.; Christiansen, O.; Hättig, C. *J. Chem. Theory Comput.* **2012**, *8*, 3274–3283.

[239] Gunaydin, L. A.; Yizhar, O.; Berndt, A.; Sohal, V. S.; Deisseroth, K.; Hegemann, P. *Nat. Neurosci.* **2010**, *13*, 387–92.

[240] Hegemann, P.; Möglich, A. *Nat. Methods* **2011**, *8*, 39–42.

[241] Tomasello, G.; Olaso-González, G.; Altoè, P.; Stenta, M.; Serrano-Andrés, L.; Merchán, M.; Orlandi, G.; Bottoni, A.; Garavelli, M. *J. Am. Chem. Soc.* **2009**, *131*, 5172–86.

[242] Welke, K.; Watanabe, H. C.; Wolter, T.; Gaus, M.; Elstner, M. *Phys. Chem. Chem. Phys.* **2013**, *15*, 6651–9.

[243] Heberle, J. *Biochim. Biophys. Acta* **2000**, *1458*, 135–147.

[244] Vogel, R.; Siebert, F. *Curr. Opin. Chem. Biol.* **2000**, *4*, 518–523.

[245] Kötting, C.; Gerwert, K. *ChemPhysChem* **2005**, *6*, 881–888.

[246] Barth, A. *Prog. Biophys. Mol. Biol.* **2000**, *74*, 141–173.

[247] Dioumaev, A. K. *Biochemistry* **2001**, *66*, 1269–1276.

[248] Dioumaev, A. K.; Braiman, M. S. *J. Am. Chem. Soc.* **1995**, *117*, 10572–10574.

[249] Braiman, M. S.; Mogi, T.; Marti, T.; Stern, L. J.; Khorana, H. G.; Rothschild, K. J. *Biochemistry* **1988**, *27*, 8516–8520.

[250] Bondar, A. N.; Smith, J. C.; Fischer, S. K. *Phase Transitions* **2005**, *78*, 671–675.

[251] Bondar, A.-N.; Smith, J. C.; Elstner, M. *Theor. Chem. Acc.* **2010**, *125*, 353–363.

[252] Radu, I.; Bamann, C.; Nack, M.; Nagel, G.; Bamberg, E.; Heberle, J. *J. Am. Chem. Soc.* **2009**, *131*, 7313–9.

[253] Schoenenberger, P.; Gerosa, D.; Oertner, T. G. *PLoS One* **2009**, *4*, e8185.

[254] Nie, B.; Stutzman, J.; Xie, A. *Biophys. J.* **2005**, *88*, 2833–2847.

[255] Yu, H.; Cui, Q. *J. Chem. Phys.* **2007**, *127*, 234504.

[256] Hanwell, M. D.; Curtis, D. E.; Lonie, D. C.; Vandermeersch, T.; Zurek, E.; Hutchison, G. R. *J. Cheminf.* **2012**, *4*, 17.

[257] Gaus, M.; Goez, A.; Elstner, M. *J. Chem. Theory Comput.* **2012**, *9*, 338–354.

[258] Weigend, F.; Ahlrichs, R. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297–305.
[259] Weigend, F. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1057–65.
[260] Adamo, C.; Barone, V. *J. Chem. Phys.* **1999**, *110*, 6158–6170.
[261] Neese, F.; Wennmohs, F.; Hansen, A.; Becker, U. *Chem. Phys.* **2009**, *356*, 98–109.
[262] Weigend, F. *J. Comput. Chem.* **2008**, *29*, 167–175.
[263] Grimme, S. *J. Chem. Phys.* **2006**, *124*, 034108.
[264] Vahtras, O.; Almlöf, J.; Feyereisen, M. *Chem. Phys. Lett.* **1993**, *213*, 514–518.
[265] Weigend, F.; Häser, M. *Theor. Chem. Acc.* **1997**, *97*, 331–340.
[266] Hellweg, A.; Hättig, C.; Höfener, S.; Klopper, W. *Theor. Chem. Acc.* **2007**, *117*, 587–597.
[267] NIST Computational Chemistry Comparison and Benchmark Database NIST Standard Reference Database Number 101 Release 15b, August 2011, Editor: Russell D. Johnson III, http://cccbdb.nist.gov/.
[268] Sander, W.; Gantenberg, M. *Spectrochim. Acta, Part A* **2005**, *62*, 902–909.
[269] Adamo, C.; Barone, V. *J. Chem. Phys.* **1998**, *108*, 664–675.
[270] Perdew, J. P. *Phys. Rev. B* **1986**, *33*, 8822–8824.
[271] Perdew, J. P. *Phys. Rev. B* **1986**, *34*, 7406–7406.
[272] Tao, J.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. *Phys. Rev. Lett.* **2003**, *91*, 146401.
[273] Grimme, S. *J. Phys. Chem. A* **2005**, *109*, 3067–3077.
[274] Schwabe, T.; Grimme, S. *Phys. Chem. Chem. Phys.* **2006**, *8*, 4398–4401.
[275] Weigend, F.; Furche, F.; Ahlrichs, R. *J. Chem. Phys.* **2003**, *119*, 12753–12762.
[276] Dunning, T. H. *J. Chem. Phys.* **1989**, *90*, 1007–1023.
[277] Kendall, R. A.; Dunning, T. H.; Harrison, R. J. *J. Chem. Phys.* **1992**, *96*, 6796–6806.
[278] Riley, K. E.; Op't Holt, B. T.; Merz, K. M. *J. Chem. Theory Comput.* **2007**, *3*, 407–433.
[279] Merrick, J. P.; Moran, D.; Radom, L. *J. Phys. Chem. A* **2007**, *111*, 11683–11700.
[280] Laury, M. L.; Carlson, M. J.; Wilson, A. K. *J. Comput. Chem.* **2012**, *33*, 2380–2387.
[281] Ireta, J.; Neugebauer, J.; Scheffler, M. *J. Phys. Chem. A* **2004**, *108*, 5692–5698.
[282] Santra, B.; Michaelides, A.; Scheffler, M. *J. Chem. Phys.* **2007**, *127*, 184104–184104–9.
[283] Zhao, Y.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *3*, 289–300.
[284] Goerigk, L.; Grimme, S. *Phys. Chem. Chem. Phys.* **2011**, *13*, 6670–6688.
[285] Biczysko, M.; Panek, P.; Scalmani, G.; Bloino, J.; Barone, V. *J. Chem. Theory Comput.* **2010**, *6*, 2115–2125.
[286] Sindhikara, D. J.; Yoshida, N.; Hirata, F. *J. Comput. Chem.* **2012**, *33*, 1536–1543.
[287] Kandori, H.; Kinoshita, N.; Shichida, Y.; Maeda, A. *J. Phys. Chem. B* **1998**, *102*, 7899–7905.
[288] Shishkin, O. V.; Gorb, L.; Luzanov, A. V.; Elstner, M.; Suhai, S.; Leszczynski, J. *J. Mol. Struct.: THEOCHEM* **2003**, *625*, 295–303.
[289] Witek, H. A.; Morokuma, K.; Stradomska, A. *J. Chem. Phys.* **2004**, *121*, 5171–8.
[290] Witek, H. A.; Morokuma, K.; Stradomska, A. *J. Theor. Comput. Chem.* **2005**, *04*, 639–655.
[291] Barone, V.; Carnimeo, I.; Scalmani, G. *J. Chem. Theory Comput.* **2013**, *9*, 2052–2071.
[292] Kaminski, S.; Gaus, M.; Phatak, P.; von Stetten, D.; Elstner, M.; Mroginski, M. A. *J. Chem. Theory Comput.* **2010**, *6*, 1240–1255.

[293] Kandori, H.; Belenky, M.; Herzfeld, J. *Biochemistry* **2002**, *41*, 6026–6031.

[294] Hayashi, S.; Ohmine, I. *J. Phys. Chem. B* **2000**, *104*, 10678–10691.

[295] Hildebrandt, P.; Stockburger, M. *Biochemistry* **1984**, *23*, 5539–5548.

[296] Martinez, M.; Gaigeot, M.-P.; Borgis, D.; Vuilleumier, R. *J. Chem. Phys.* **2006**, *125*, 144106–144106–14.

[297] Baer, M.; Mathias, G.; Kuo, I.-F. W.; Tobias, D. J.; Mundy, C. J.; Marx, D. *ChemPhysChem* **2008**, *9*, 2703–7.

[298] Ebnet, E. *Plant Cell* **1999**, *11*, 1473–1484.

[299] Prochnik, S. E. et al. *Science* **2010**, *329*, 223–6.

[300] Pfeiffer, F.; Schuster, S.; Broicher, A.; Falb, M.; Palm, P.; Rodewald, K.; Ruepp, A.; Soppa, J.; Tittor, J.; Oesterhelt, D. *Genomics* **2008**, *91*, 335–346.

[301] Otomo, J.; Muramatsu, T. *Biochim. Biophys. Acta, Biomembr.* **1995**, *1240*, 248–256.

[302] Vogeley, L.; Sineshchekov, O. A.; Trivedi, V. D.; Sasaki, J.; Spudich, J. L.; Luecke, H. *Science* **2004**, *306*, 1390–3.

[303] Tsunoda, S. P.; Ewers, D.; Gazzarrini, S.; Moroni, A.; Gradmann, D.; Hegemann, P. *Biophys. J.* **2006**, *91*, 1471–9.

[304] Sineshchekov, O. A.; Govorunova, E. G.; Jung, K.-H.; Zauner, S.; Maier, U.-G.; Spudich, J. L. *Biophys. J.* **2005**, *89*, 4310–9.

[305] Gradinaru, V.; Zhang, F.; Ramakrishnan, C.; Mattis, J.; Prakash, R.; Diester, I.; Goshen, I.; Thompson, K. R.; Deisseroth, K. *Cell* **2010**, *141*, 154–65.

[306] Espagne, E. et al. *Genome Biol.* **2008**, *9*, R77.

[307] Idnurm, A.; Howlett, B. J. *Genome* **2001**, *44*, 167–71.

[308] Bieszke, J. A.; Braun, E. L.; Bean, L. E.; Kang, S.; Natvig, D. O.; Borkovich, K. A. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 8034–9.

[309] Tateno, M.; Ihara, K.; Mukohata, Y. *Arch. Biochem. Biophys.* **1994**, *315*, 127–32.

[310] Ihara, K.; Umemura, T.; Katagiri, I.; Kitajima-Ihara, T.; Sugiyama, Y.; Kimura, Y.; Mukohata, Y. *J. Mol. Biol.* **1999**, *285*, 163–74.

[311] Baliga, N. S.; Bonneau, R.; Facciotti, M. T.; Pan, M.; Glusman, G.; Deutsch, E. W.; Shannon, P.; Chiu, Y.; Weng, R. S.; Gan, R. R.; Hung, P.; Date, S. V.; Marcotte, E.; Hood, L.; Ng, W. V. *Genome Res.* **2004**, *14*, 2221–34.

[312] Frassanito, A. M.; Barsanti, L.; Passarelli, V.; Evangelista, V.; Gualtieri, P. *Cell. Mol. Life Sci.* **2010**, *67*, 965–71.

[313] Inoue, K.; Ono, H.; Abe-Yoshizumi, R.; Yoshizawa, S.; Ito, H.; Kogure, K.; Kandori, H. *Nat. Commun.* **2013**, *4*, 1678.