Influence of Protein and Solvent Environments on Quantum Chemical Properties of Photosynthesis Enzymes and Photoreceptors

Dissertation

zur Erlangung des akademischen Grades

“doctor rerum naturalium”

(Dr. rer. nat.)

in der Wissenschaftsdisziplin Biochemie

eingereicht an der

Mathematisch-Naturwissenschaftlichen Fakultät

der Universität Potsdam

von

Jan Philipp Götze

aus Berlin

Potsdam, 15. September 2010

Published online at the
Institutional Repository of the University of Potsdam:
URL http://opus.kobv.de/ubp/volltexte/2011/5113/
URN urn:nbn:de:kobv:517-opus-51135
http://nbn-resolving.org/urn:nbn:de:kobv:517-opus-51135

“Jan, sagen Sie mal, hat das überhaupt noch etwas mit Chemie zu tun?”

Dipl.-Chem. Irmhild Steinke (1940 - 2007)

bei der Habilitationsfeier von Dr. Tillmann Klamroth

Zusammenfassung

In dieser Arbeit wird an drei Beispielen gezeigt, inwiefern sich die unterschiedliche Beschreibung der Umgebung auf die Ergebnisse der Berechnung biologischer Prozesse auswirkt. Hierzu wird die Umgebung, entweder Lösungsmittel oder Protein, manchmal explizit in die Berechnung des Systems integriert, oder manchmal auch nur näherungsweise in die Berechnung mit aufgenommen.

Für das erste System, das Enzym Ribulose-1,5-bisphosphat Carboxylase/Oxygenase, wird ein kleiner Teil des aktiven Zentrums in verschiedenen Varianten simuliert. Ein bestimmter Reaktionsschritt, die Carboxylierung von Ribulose-1,5-bisphosphat, wird hierbei auf seine Massendiskriminierung bezüglich eines Kohlenstoffatoms hin untersucht.
Hierzu wurde der Reaktionsschritt mit Edukten, Produkten und Übergangszustand modelliert und energetische Profile für die Reaktionspfade erstellt.
Es hat sich gezeigt, dass sich die verschiedenen Geometrien des aktiven Zentrums minimal auf den Isotopeneffekt auswirken. Die Umgebung scheint für diese Eigenschaft also nur eine untergeordnete Rolle zu spielen.

Das zweite System, die Blaulicht-Rezeptordomäne BLUF (blue-light sensing using flavin), wurde in Bezug auf die Funktionalität der umgebenden Aminosäuren hin untersucht.
Hierzu wurden UV/vis Spektren eines NMR basierten Strukturensembles berechnet. Da die ersten Ergebnisse zeigten, dass eine bisher unbeschriebene Seringruppe an dem spektralen Verhalten der BLUF Domäne beteiligt sein könnte, wurden weitere Untersuchungen angeschlossen. Diese bestanden aus Molekulardynamik bei Raumtemperatur verschiedener BLUF Domänen und die daraus resultierenden Wasserstoffbrückenbindungen, neben erneuter Berechnung der UV/vis Spektren.
Die Bindungseigenschaften der Seringruppe waren stark von der globalen Konformation des Proteins abhängig. Da wir gezeigt haben, dass die Seringruppe das UV/vis Spektrum beeinflussen kann, erwarten wir durch die verschiedenen globalen BLUF Konformationen eine Änderung des UV/vis Spektrums, was im Rahmen unserer Studie jedoch nicht gezeigt werden konnte. Experimentelle Daten unterstützen jedoch inzwischen Teile unserer Berechnungen bzgl. der Seringruppe.

Schließlich wurde auch ein Teil des Lichtsammelkomplex II (LHC II) untersucht, speziell das Chromophorenpaar Violaxanthin/Chlorophyll b.
Es wurden sowohl Polarisationsmodelle für die Berechnung der Eigenschaften in Lösung verwendet, als auch Punktladungsfelder für die Nachahmung des gesamten Proteins.
Es stellte sich heraus, dass das Absorptionsspektrum von Violaxanthin in LHCII im Verhältnis zum Spektrum in Lösung stark rot verschoben sein sollte. Dieses Ergebnis zeigte sich unabhängig von den benutzen Rechenmethoden oder Umgebungen. In manchen Fällen war der Abstand zwischen den beiden niedrigsten Absorptionsbanden des Chromophorenpaars so klein, dass sich die Möglichkeit für eine neue Art von Lichtsammelmechanismus in LHCII aufgetan hat. Dieser sollte in zukünftigen Studien näher untersucht werden.

Diese Arbeit zeigt, wie wichtig oder wie unwichtig die Einbeziehung der Umgebung für bestimmte biologische Fragestellungen sein kann.

Copyright

Acknowledgements

I have to admit that this section of my thesis was by far the hardest to assemble. Especially the high diversity of the topics of this work made it difficult, since most of the splendid people I would like to thank were involved in only one of the projects.

This is however not the case for most of the *Theoretical Chemistry group* at the University of Potsdam. Some of them would have probably liked to be less involved. Yet, I think *Prof. Peter Saalfrank* is happy to, as he seemed to become more and more interested in biochemical topics. The amount of corresponding books in his shelf indicated as much. Thank you for your patience, advice and, of course, the money.
Speaking of patience, *PD Dr. Tillmann Klamroth* should be mentioned. I have to thank him for the advice on technical matters and his constant scolding because of my sloppy techniques (which became less and less sloppy thanks to him). Last but not least, of the Potsdam Postdocs, *Dr. Dominik Kröner* deserves praises for performing such fabulous work on the LHCII system (and, probably, on the Wii system, too). He agreed to examine this thesis carefully during the process of writing, for which I am most grateful. *Bastian Klaumünzer*, although not having contributed directly to this thesis, always was my "other side of the medal", as he was responsible for the IR and Raman of the BLUF project, as well as for the excited state calculations. Without all of you, this work would simply not exist.

Having started mentioning the people working in a specific project, I have to thank *Dr. Heiko Lokstein* (UP) for the advice on LHC II. In case of the RuBisCO project, I want to thank *Dr. Oliver Ebenhöh* (Aberdeen) and *Franziska Witzel* (Charite Berlin). Finally, regarding the BLUF project, there are too many people to mention them all, but *Dr. Claudio Greco* (Milan) and *Dr. Tilo Mathes* (HU Berlin) definitely deserve my greatest gratitude.

I would also like to thank everybody who agreed to be a reviewer for this thesis. I hope you will enjoy reading it and it will live up to your expectations.

Finally, my office mates deserve to be mentioned, with who I shared D.1.19-20 in Golm for several months or even years during my thesis. In chronological order, Dr. Tijo Vazhappily (now U.S.), Dr. Padmanaban Ramanathan (now back in India), Stefan Klinkusch (still there), Gereon Floß (still there) and Monique Mertens (still there). I always enjoyed the time with you and I hope you will keep up the good work.

Contents

1 Introduction **1**

1.1 The Context of Theory and Biology 1

1.2 Computational Modeling of Photosynthesis Proteins 3

1.2.1 Photosynthesis 4

1.2.2 Investigated Systems 5

2 Methodology **9**

2.1 Quantum chemistry 9

2.1.1 Basic concepts 10

2.1.2 Hartree-Fock and derived methods 12

2.1.3 Density functional methods 17

2.1.4 Calculation of UV/vis spectra 21

2.1.5 Normal Mode Analysis and Thermal Correction 24

2.1.6 Transition states 26

2.1.7 Polarizable Continuum Model 27

2.1.8 Electronic embedding 29

2.2 Force Field Theory and Application 30

2.2.1 Force Fields 31

2.2.2 Molecular Dynamics 33

2.3 QM/MM methods 37

2.3.1 Additive Approach 37

2.3.2 Subtractive Approach 38

3 The RuBisCO Enzyme **41**

3.1 The RuBisCO carbon isotope effect 41

3.1.1 RuBisCO: Function and abundance 41

3.1.2 Active site of RuBisCO . . . 42
3.1.3 Carboxylation Step and Model construction . . . 45
3.2 Models of the Carboxylation Step . . . 46
3.2.1 Conventional Model . . . 46
3.2.2 New Model . . . 49
3.3 Discussion and Summary (RuBisCO) . . . 50

4 BLUF domain characterization 55

4.1 Blue-light receptor domains . . . 55
4.1.1 BLUF domain properties . . . 56
4.1.2 Structure/function assignment . . . 58
4.1.3 Roles of single amino acids . . . 61
4.2 Spectral characterization . . . 62
4.2.1 Model construction . . . 63
4.2.2 Spectral analysis . . . 64
4.2.3 Structural analysis . . . 67
4.3 Effects of global BLUF conformation . . . 71
4.3.1 Model construction . . . 72
4.3.2 MD protocol . . . 73
4.3.3 Hydrogen bonding . . . 74
4.3.4 MD snapshot based UV/vis spectra . . . 76
4.4 Summary and Discussion (BLUF) . . . 79

5 The LHCII complex 83

5.1 The LHCII energy transfer . . . 83
5.1.1 Structure and Function of LHCII*b* . . . 84
5.1.2 Light harvesting . . . 85
5.2 Modeling of Initial Light Absorption in Pea LHCII . . . 87
5.2.1 Model Construction . . . 87
5.2.2 Isolated Molecules . . . 89
5.2.3 Coupled chromophore systems . . . 93
5.3 Summary and Discussion (LHCII) . . . 97

6 Conclusions 101

6.1 Summary of the Thesis 101
6.2 Future Perspectives 103

A **121**

A.1 Details on RuBisCO Results 121
A.1.1 6-31G(d) Test Calculations for the New Model 121
A.1.2 ZPE values 122
A.2 Details on LHCII Results 123
A.2.1 Other LHCII subunits 123
A.2.2 Effect of Supporting Amino Acids 124
A.2.3 Effect of the Phytyl Residue 125
A.2.4 Effect of His Charge 126
A.2.5 Effect of Basis Set 127
A.2.6 CIS results 128

Chapter 1

Introduction

1.1 The Context of Theory and Biology

In the last two decades, computer-based, theoretical research on biological systems has increased dramatically. This is due to the increasing desire to understand a biological system either in exact molecular detail and/or as an interconnected complex system. This is complemented by increasing computational resources, as well as advances in computational algorithms. The resulting research field splits mainly into two branches: First, pathway or network modeling, an important portion of so-called "Systems Biology". This is a very promising approach to combine experimental research and theoretical modeling. However, pathway modeling in general does not require any physical basis. An abstract reaction network can be build, extended or modified without taking the actual molecular construction of the metabolites or enzymes into account [1, 2, 3].

The second branch, which incorporates the molecular background, is represented by chemical simulations of biological systems. Since the computers have become more and more powerful, as well as the theoretical description of molecules more and more efficient, biological systems can nowadays be addressed with good accuracy. In principle, chemical simulations try to grasp the events of (bio-)chemistry on a molecular basis. In the ideal case, a computational chemist will end up with one or many "movies" of the process, from which many chemical and physical properties can be deduced. Enthalpies of formation, activation barriers and specific molecular interactions are of primary interest. More sophisticated analysis is possible, too, depending on the accuracy of the systems' description [4, 5, 6].

The methods used for these simulations were at first exclusively based on wave functions describing the system (*ab initio* methods [7, 8]) or density functional theory [9]. First steps to conduct quantum chemical research on biologically relevant molecules reach back to at least 1958 [10]. While calculations using these quantum chemical methods can be highly accurate, they cannot be used for systems larger than a few hundred atoms. In addition, standard quantum mechanics tend to be restricted to idealized systems *in vacuo* at 0 K. There are ways to overcome these limits, but especially the grasp of combined nuclear movement of all atoms is difficult on the level of quantum chemistry.
Biological systems, due to their large number of atoms moving at ambient temperatures, will still not be described on a full quantum mechanical level for at least several decades. The amount of electron-electron interaction terms exceeds the computational resources, even using very simple approaches. Most scientists wish for a calculation to complete at least within their lifetime, for obvious reasons. Thus, large systems of more than about a hundred to several thousand atoms require a different computational treatment: Force fields are generally applied to such problems, instead of explicit description on the electronic level [11, 12, 13, 14].
Force fields treat the whole system purely on the level empirical or previously calculated potentials. To do so, various empirical parameters have to be defined beforehand. Force field-based dynamics (molecular dynamics, MD) enables the modeler to observe movement at a given temperature, system evolution in time, and other properties not easily available from quantum chemical calculations. Yet, the system's properties are somewhat artificial, as any event involving the change of electron distribution is simply neglected. In standard force fields, no bonds can be broken, induced dipoles cannot be observed, and so on. We will see later on that there are ways to avoid some of these artifacts.

With the toolkit at hand, the problem of system borders arises. Commonly, one will define an active region, like the active site of an enzyme or the chromophores in a photoreceptor. Restriction to only these parts may however lead to severe artifacts. The surrounding protein structure, mostly referred to as "environment" is most of the time directly or indirectly coupled to the active region. Simply omitting this environment may consequently change the quality of the events in the active region, which has to be avoided

to obtain realistic results.
In this context one has to consider not only the presence, but also the complexity of the environment. Experimental findings do not always lead to a clear picture of the events in an active region. Details in active site geometry may lead to different reaction pathways. A central aspect of this thesis is therefore to investigate the effects of changing the exact reaction mechanisms or system borders for different biological problems. We will see that some details do not lead to any qualitative changes, while others have tremendous influence on the picture of a biological process.

This thesis will focus on the description of a few selected biological systems, which are all connected to processes in photosynthesis or regulation thereof. Both the presented force field and the quantum chemical approaches will be used. We will also see that in some cases, results from the one level can be used for calculations on the other, and *vice versa*. This creates a bridge between the quantum chemistry and force fields, without the need of explicit coupling [15]. More on this later on throughout this thesis. For now, we will focus on the general challenges of computational chemistry in a biological context in greater detail.

1.2 Computational Modeling of Photosynthesis Proteins

The aim of this section is to present the general context as well as the biological systems used for our purposes. All biological systems addressed in this thesis are protein based. The properties of a complex protein environment result in at least three problems:

- A quantum chemical model is not capable of including "all" atoms of a large biological system, as computational effort scales extensively with the amount of electrons. Thus, long-range and solvent effects have to be omitted.

- The active region has to be cut from the system at some point creating a border. This border may induce geometrical distortions, shifts in electron distribution or other artifacts.

- Dynamic effects, such as protein movement at room temperature, is not covered in static quantum models. Sometimes, if structural data from X-ray and NMR are not used unoptimized, the optimized structures correspond to 0 Kelvin.

We explore these problems in detail throughout this thesis. To do so, we employ three different model systems, which all have a direct connection to photosynthesis.

1.2.1 Photosynthesis

Photosynthesis in the core process for sustaining life on earth. Photosynthesis takes place in cyanobacteria and in the chloroplast organelles of algae and plants. For these organisms, photosynthesis provides metabolite precursors and energy for biomass production. All other organisms on earth depend on this biomass for nutrition, and many also on the byproduct oxygen, too. It is therefore save to say that photosynthesis is one of the most extensively studied biochemical processes and will stay so in the years to come.
Photosynthesis is split in two general reaction complexes, the *light reactions* and the *dark reactions*. The former are dependent on photon energy (hence the name). They consist of mainly electron and ion transport processes along or through the thylakoid membrane. The purpose of the light reaction is to provide energy in the form of adenosine triphosphate (ATP) and high-energy electrons for reduction processes in the form of reduced nicotinamide adenine dinucleotide phosphate (NADPH). As source of the electrons required for the latter, water is split into H^+ ions and oxygen. The ions are used for the generation of ATP and O_2 is formed and released as "waste" product.

Dark reactions is another term for the Calvin-Benson cycle. It is a CO_2 consuming cyclic reaction pathway featuring exclusively sugars with a length ranging from 3 to 7 carbon atoms. The purpose of the cycle is to produce 3-carbon sugar phosphates (triose phosphates), which are exported to the cell plasma in case of algae or plants. To make up for the loss of phosphate in the transport process, the process is facilitated by a phosphate/triose phosphate anti-transporter. As only one CO_2 is assimilated in each cycle, the products are used most of the time to rebuild the 5-carbon sugars required for the carbon fixation step.

The Calvin-Benson cycle is the source for almost all organic carbon in living organisms. The cycle is fueled by the ATP and NADPH produced in the light reactions. The metabolic coupling between the light and dark reactions is primarily spatial, as they occur in the same organelle. However, regulatory processes may affect both the dark and light reactions, lowering quantum yield in excess light, or lowering Calvin cycle activity in the dark [16, 17].

This thesis will focus on events at active regions of several proteins participating in reaction of photosynthesis or regulation thereof. For an overview, the processes of photosynthesis are depicted in fig. 1.1. We will now briefly present the treated systems in their biological context. Note that for each of the model systems, there will be a separate, detailed introduction in the beginning of chapters 3, 4, and 5.

1.2.2 Investigated Systems

First, we explore the carbon isotope effect of ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) [18] in chapter 3. RuBisCO is the key player in the Calvin-Benson cycle, as it facilitates the CO_2 fixation step. The goal is to explore the origin of a $^{12}C/^{13}C$ isotopic effect and its dependence on the proposed mechanism. Model size and environmental influences are briefly discussed, too. This is done by setting up various model systems and calculation with various computational methods.
Second, we present our results on the spectral characterization of a blue light sensor domain (blue-light using flavin, BLUF [19]) in chapter 4. Some proteins containing BLUF domains are known to be capable of regulating the light reactions of photosynthesis. Our aim is to find new possible roles for amino acids in the active region. In contrast to the RuBisCO system, the definition of the active region is mainly invariant with respect to the atoms included, but structural dynamics are investigated as well. The quantum chemical methodology is also simpler than in the RuBisCO case. However, the protein environment is taken into account in form of a field of Coulomb point charges.
Finally, in a similar manner, we study the properties of the light harvesting complex II (LHCII [16]) in chapter 5. The LHCII is a chromophore carrier, which is indirectly

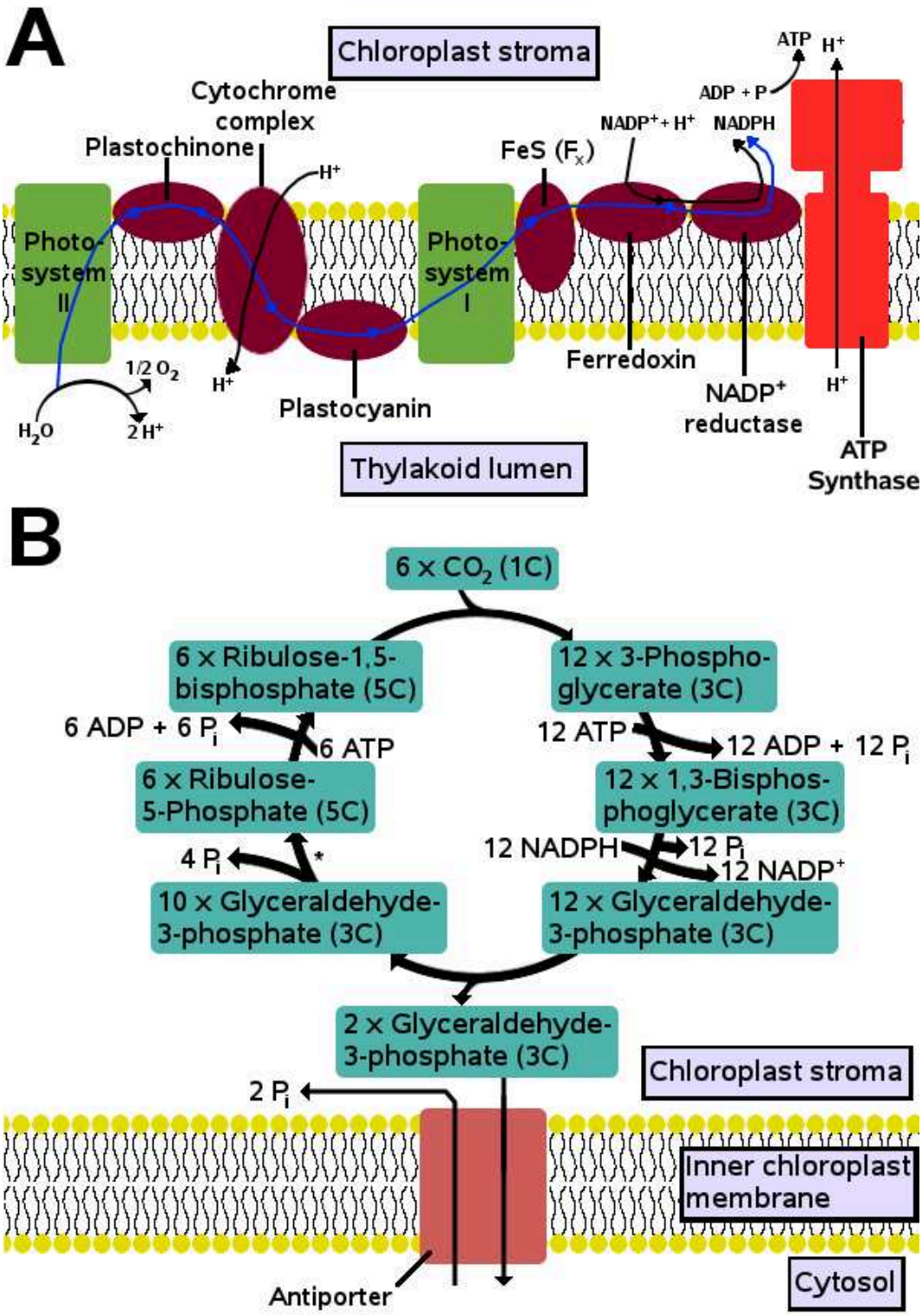

Figure 1.1: Overview on photosynthesis processes. **A**: Light reactions (blue arrows inside the thylakoid membrane: electron transport chain). Photosystems II and I require photon energy to excite electrons to high-energy states, which can enter the electron transport chain thereafter. **B**: Dark reactions (Calvin-Benson cycle). The metabolic cycle processes carbon atoms to yield glyceraldehyde-3-phosphate. Several enzymatic steps summarized in the arrow marked with * (regeneration phase). For the different phases of the cycle, see fig. 3.2.

coupled to the photosystem II (see fig. 1.1A). It is capable of supporting the harvesting of light energy. It is also capable of showing a mechanism of photoprotection known as non-photochemical quenching (NPQ) [16]. Here, we put more emphasis on methodology and environmental effects in contrast to the BLUF case, as the mechanism is widely acknowledged as well understood.

All three systems share the problem of an appropriate definition of model borders and of how to treat a complex environment. This becomes a challenge regarding the approaches required to treat these systems. In the next chapter, we will consequently discuss the theory used for the analysis of the presented model systems. We will see that using a combination of various methods helps to overcome the general problems for the simulation of biological systems.

Chapter 2

Methodology

In this chapter, the computational methods used in this thesis will be presented. The supporting concepts (*e.g.*, basis sets) will be discussed when needed. We will start with standard quantum chemical methods (sec. 2.1). After an overview on classical methods (sec. 2.2), the coupling schemes for both quantum and classical methods will be displayed (sec. 2.3). Most of this section's content was acquired from textbooks [20, 21].

2.1 Quantum chemistry

One purpose of quantum chemistry (QC) or quantum mechanics (QM), is to calculate the energy and properties of a molecule. To address small, isolated molecular systems, QM is generally the most successful approach. Unfortunately, biological systems like proteins in aqueous solution are rarely small and never isolated.

As such, improvement of the computational methods towards the treatment of larger systems is and has always been one of the central aims in theoretical research since the middle of 20th century. QM calculations result in a multi-body problem of rapidly increasing complexity. Therefore, much research in the last decades aimed at reducing the scaling of computation time with system size. As a consequence, the original, electronic wave function based Hartree-Fock methods (HF [7, 8], sec. 2.1.2) are nowadays less used than density functional theory (DFT [9], sec. 2.1.3) for larger molecules. This is due to their different scaling with the number of electrons, which is rapidly rising for more advanced wave function methods. However, DFT has the disadvantage of not being systematically improvable, *e.g.*, the molecule might be predicted to be too stable. The HF methods will always yield a ground state energy which is higher or equal to the correct

value.

The correct choice of computational method (time vs. accuracy) is one of the main skills of a quantum chemist. In this work, several QM methods were used, which will be presented in greater detail in this section. An additional direct improvement to these methods is the polarizable continuum approach, introducing solvent effects to the simulation without taking each solvent molecule explicitly into account. This is further discussed in sec. 2.1.7.

As already indicated above, larger molecular systems cannot be treated efficiently with QM methods, the time required to finish the calculation and also the sheer amount of required memory is often too large. Even semi-empirical methods, which strongly rely on empirical parameters for the most time-consuming calculation steps (the two-electron integrals, see sec. 2.1.2), will face their limits in the range of several hundred atoms. To overcome this problem, force fields have been developed, which will be discussed in sec. 2.2.

2.1.1 Basic concepts

Arising from the quantization of energy levels at small dimensions, the energy (or stability) of any molecular system is not exactly tractable with classical mechanics. As a result, wave mechanics or quantum mechanics were developed, which give rise to the time-dependent Schrödinger equation

$$i\hbar\frac{\partial\Psi}{\partial t} = \hat{H}\Psi \tag{2.1}$$

with i^2 being -1, $\hbar$ Planck's constant divided by 2π, Ψ being the wave function describing the current state of the system, t being the time and, finally, $\hat{H}$ being the Hamiltonian of the system, yielding the system's energy if applied to Ψ. The time-dependent Schrödinger equation is the basis for quantum dynamics and quantum mechanics alike. For time-independent problems (2.1) simplifies to

$$\hat{H}\Psi = E\Psi \tag{2.2}$$

which is the time-independent Schrödinger equation. In principle, it is possible to extract the energy E of every system from the wave function Ψ, as the Hamiltonian $\hat{H}$ is simple to construct.

The computational problem arises from Ψ, which is not analytically given for complex systems (meaning more than one electron). Ψ can however be simplified, separating the treatment of atomic nuclei and electrons. This is called the Born-Oppenheimer approximation.

The Born-Oppenheimer separation

A molecule consists of N electrons and N_A nuclei. Assuming the nuclear motion to be much slower than the motion of the electrons, it can be stated that the electrons will react almost instantaneously to any change of the nuclear coordinates. This assumption can be made from the fact that the mass m_A of any nucleus A, even a single proton, is much larger than the electron mass m_e.

Furthermore, treating the nuclei as immobile point charges within the electronic problem allows for treating the nuclear repulsion as a simple Coulomb potential (see below). Thus, the wave function Ψ of the system can be written as a product of the *electronic* wave function Ψ_{el} and the *nuclear* wave function Ψ_{nuc}. Ψ_{el} then depends only on electronic coordinates $\underline{\mathbf{r}}$; the nuclear coordinates $\underline{\mathbf{R}}$ are treated as parameters:[1]

$$\Psi(\underline{\mathbf{r}}, \underline{\mathbf{R}}) = \Psi_{el}(\underline{\mathbf{r}}; \underline{\mathbf{R}}) \cdot \Psi_{nuc}(\underline{\mathbf{R}}) \quad . \tag{2.3}$$

The *electronic* Schrödinger equation is thus given by

$$\hat{H}_{el}\Psi_{el}(\underline{\mathbf{r}}; \underline{\mathbf{R}}) = E_{el}\Psi_{el}(\underline{\mathbf{r}}; \underline{\mathbf{R}}) \quad . \tag{2.4}$$

$\hat{H}_{el}$ is the electronic Hamiltonian without the nuclear repulsion and nuclear kinetic energy (in contrast to $\hat{H}$ from (2.2)). The electronic Hamiltonian reads

$$\begin{aligned} \hat{H}_{el} &= \hat{T}_{el}(\underline{\mathbf{r}}) + \hat{V}_{el-nuc}(\underline{\mathbf{r}}, \underline{\mathbf{R}}) + \hat{V}_{el-el}(\underline{\mathbf{r}}) \\ &= -\frac{\hbar^2}{2m_e}\sum_i^N \nabla_i^2 - \sum_i^N \sum_A^{N_A} \frac{Z_A e^2}{4\pi\epsilon_0 |\mathbf{r}_i - \mathbf{R}_A|} + \sum_i^{N-1} \sum_{j>i}^{N} \frac{e^2}{4\pi\epsilon_0 |\mathbf{r}_i - \mathbf{r}_j|} \end{aligned} \tag{2.5}$$

with $\hat{T}_{el}$ as the kinetic energy operator for the electrons, $\hat{V}_{el-nuc}$ as attractive potential between the nuclei and the electrons and finally $\hat{V}_{el-el}$ as the electron-electron repulsion.

[1] Note that $\mathbf{r}$ is in bold face, representing the set of three spatial coordinates for each electron. The same is true for the nuclear coordinates $\mathbf{R}$. The set of all electronic or nuclear coordinates, $\underline{\mathbf{r}}$ and $\underline{\mathbf{R}}$, respectively, therefore are vectors having $3N$ or $3N_A$ elements.

It is useful to introduce atomic units at this point ($\hbar = m_e = e = 4\pi\epsilon_0 = 1$), as (2.5) then reads as

$$\hat{H}_{el} = -\frac{1}{2}\sum_{i}^{N}\nabla_i^2 - \sum_{i}^{N}\sum_{A}^{N_A}\frac{Z_A}{|\mathbf{r}_i - \mathbf{R}_A|} + \sum_{i}^{N-1}\sum_{j>i}^{N}\frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} \tag{2.6}$$

We will use atomic units throughout the rest of this work.

From the construction of $\hat{H}_{el}$ in (2.5) and (2.6), it is clear that the resulting electronic energy E_{el} in (2.4) is not the complete energy of the system: It misses the nuclear repulsion, if the nuclei are considered immobile ($\left\langle \hat{T}_{nuc} \right\rangle = 0$). For a given set of nuclear coordinates, a geometry, the energy of the molecule must be written as

$$E = E_{el} + V_{nuc-nuc}(\underline{\mathbf{R}}) \tag{2.7}$$

with

$$V_{nuc-nuc}(\underline{\mathbf{R}}) = \sum_{A}^{N_A-1}\sum_{B>A}^{N_A}\frac{Z_A Z_B}{|\mathbf{R}_A - \mathbf{R}_B|} \quad . \tag{2.8}$$

The Born-Oppenheimer separation allows for scans of potential energy curves. These are sets of several fixed nuclear coordinates, connected by the change of nuclear coordinates. From there, energy barriers and reaction paths can be derived. Combinations of multiple nuclear reaction coordinates lead to energy (hyper-)surfaces of higher dimensionality (2D, 3D and so on).

To find an approximate wave function for the system, we can make use of several wave function based methods, which will be presented in the next section.

2.1.2 Hartree-Fock and derived methods

We need to calculate a numerical electronic wave function $\Psi_{el}(\underline{\mathbf{r}}; \underline{\mathbf{R}})$ for a given N and a fixed nuclear geometry. A simple product of N single electron wave functions (Hartree product) does not fulfill Pauli's principle. To obtain an acceptable ground state electronic wave function Ψ_{el} for a given molecular geometry described by $\hat{H}_{el}$, it is useful to denote Ψ_{el} as a (normalized) Slater determinant

$$\Psi_{el}(\underline{\mathbf{x}}) = \frac{1}{\sqrt{N!}} \cdot \begin{vmatrix} \chi_1(\mathbf{x}_1) & \chi_2(\mathbf{x}_1) & \cdots & \chi_N(\mathbf{x}_1) \\ \chi_1(\mathbf{x}_2) & \chi_2(\mathbf{x}_2) & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \chi_1(\mathbf{x}_N) & \cdots & \cdots & \chi_N(\mathbf{x}_N) \end{vmatrix} \tag{2.9}$$

with the orthonormal spin orbitals

$$\chi(\mathbf{x}_i) = \psi(\mathbf{r}_i)\gamma(\omega_i) \quad . \tag{2.10}$$

Here, $\psi(\mathbf{r})$ denotes a spatial orbital and $\gamma(\omega)$ a spin function with $\alpha(+1/2) = 1$ or $\beta(-1/2) = 1$, and zero in all other cases. In (2.9), the spatial coordinates $\mathbf{r}$ and the spin coordinate ω have been combined in an electronic coordinate $\mathbf{x}$. Note that writing Ψ_{el} in the form of (2.9) already fulfills Pauli's principle of antisymmetry upon exchange of two electrons.

We now require the optimal spin orbitals $\chi(\mathbf{x})$ which yield the lowest possible electronic energy of the system via (2.4).

This is due to the *variational principle for wave functions*, which states that the ground state energy of the system is equal to the lowest possible energy, and thus lower than or equal to the energy of all other possible (normalized) trial functions $\tilde{\Psi}_{el}$:

$$E_0 \leq \int \tilde{\Psi}_{el} \hat{H}_{el} \tilde{\Psi}_{el} d\underline{\mathbf{x}} \tag{2.11}$$

No $\tilde{\Psi}_{el}$ will yield the exact ground state energy of the system, as we neglected relativistic effects and nuclear motion. We assumed as well that a Hartree product in form of a single Slater determinant is enough for the description of the electronic wave function, which yields an additional error (correlation energy, see other section(s)).

Hartree-Fock (HF)

The step-by-step derivation of the Hartree-Fock scheme [7, 8] would exceed the limits of this thesis, and is therefore omitted. The basic idea arising from (2.11) is

$$\frac{\delta E}{\delta \chi_i} = 0 \quad \forall \quad \chi_i \tag{2.12}$$

The resulting Hartree-Fock (HF) equations are of the form

$$\hat{f}\chi_i(\mathbf{x}_1) = \epsilon_i \chi_i(\mathbf{x}_1) \tag{2.13}$$

with ϵ_i as the orbital energy. The Fock operator $\hat{f}$

$$\hat{f}(\mathbf{x}_1) = \hat{h}(\mathbf{x}_1) + \sum_{l}^{N} \left[\hat{J}_l(\mathbf{x}_1) - \hat{K}_l(\mathbf{x}_1)\right] \tag{2.14}$$

with $\hat{h}_i$ as one-electron Hamiltonian. The Coulomb operator $\hat{J}_l$ accounts for electron-electron repulsion. The exchange operator $\hat{K}_l$ corrects the Coulomb-based electron-electron repulsion in terms of spin correlation, resulting in a stabilization if two electrons possess the same spin.

In this thesis, for the wave function based approaches, exclusively *restricted* Hartree-Fock (RHF) is used. It can be shown that for the calculation of most properties, the spin function γ can be eliminated. Two electrons only differing in the spin can therefore be considered occupying the same spatial orbital $\psi(\mathbf{r})$. This yields the form of the RHF wave function for which we only have to find $N/2$ spatial orbitals instead of N spin orbitals.

Basis sets

Technically, the HF equations (2.13) are coupled single-electron Schrödinger equations (and exact in the case of a one-electron system). They can be solved if the spatial part of the molecular orbitals (MO) is represented as a linear combination of atomic orbitals (AO), the LCAO-MO approach:

$$\psi_i(\mathbf{r}) = \sum_{j}^{N_{AO}} c_{ji}\phi_j(\mathbf{r}) \tag{2.15}$$

The AOs themselves are represented in form of a basis set. All basis sets used in this thesis are of the Pople type. The occupied orbitals of an atom are represented in terms of contracted Gaussian functions (contracted Gaussian type orbitals, CGTO). At least one of the CGTO functions is used for each considered atomic orbital on each atom. For certain applications, one may choose to include unoccupied orbitals as well (see below).

In general, a PGTO (P for "primitive") in Cartesian coordinates is represented by

$$\phi_{\zeta,l_x,l_y,l_z}(x,y,z) = N_n x^{l_x} y^{l_y} z^{l_z} e^{-\zeta r^2} \tag{2.16}$$

with $r = \sqrt{x^2 + y^2 + z^2}$ and N_n as normalization factor. The sum of all l exponents yields the orbital type (s, p, d etc., depending on the number of radial nodes). To obtain a more realistic function (*e.g.*, asymptotic behavior), k PGTOs are usually summed up to a single CGTO (C for "contracted"):

$$\phi(CGTO) = \sum_{i}^{k} a_i \phi_i(PGTO) \tag{2.17}$$

Furthermore, Pople-style basis sets allow for further refinement of the valence orbitals via *split valence* functions. Split valence indicates the usage of multiple basis functions for the same atomic orbital. It is useful as it allows for more flexibility in the valence shell. It is also possible to introduce atomic orbitals of higher, unoccupied shells like 2p-functions for hydrogen or 3d-functions for second row elements. These orbitals enable for, *e.g.*, a more sophisticated description of atoms bound to aromatic systems or the description of an asymmetric electron density within a bond ("polarization"). It is due to the latter that these functions are called polarization functions.

Used for most of the calculations in this thesis were split valence basis sets with polarization functions ("6-31G(d)" or "6-31G(d,p)"; 6 PGTOs in a CGTO for non-valence shells, for valence shells two CGTOs made from 3 or 1 PGTOs, respectively; (d) and (d,p): second row elements supported by d-functions, and hydrogen supported by p-functions).

Roothaan-Hall equations

Once a basis of M basis functions is introduced, according to (2.15) we do not have to calculate the wave functions themselves. Instead, we need to find the coefficients c_{ji} for each of the basis functions ϕ_j for the ith orbital. This can be done using the Roothaan-Hall equations

$$\underline{\underline{F}}\ \underline{c} = \underline{\underline{S}}\ \underline{c}\ \underline{\epsilon} \tag{2.18}$$

with the Fock matrix

$$F_{ij} = \int \phi_i^*(\mathbf{x}_1) \hat{f}(\mathbf{x}_1) \phi_j(\mathbf{x}_1) d\mathbf{x}_1 \tag{2.19}$$

and the overlap matrix

$$S_{ij} = \int \phi_i^*(\mathbf{x}_1)\phi_j(\mathbf{x}_1)d\mathbf{x}_1 \tag{2.20}$$

(2.18) can be easily derived from (2.13) and (2.15) using above definitions. By guessing a starting trial set of coefficients, we can solve (2.18) to obtain new coefficients and updated ϵ values. The updating is repeated until convergence according to (2.12) is achieved. Each iteration is called a SCF (self-consistent field) cycle.

Semi-empirical methods (PM6)

The Fock matrix elements F_{ij} can become computationally expensive, as any element is explicitly written as, in the RHF case,

$$F_{ij} = h_{ij} + \sum_{u}^{N/2}\sum_{l}^{N_{AO}}\sum_{m}^{N_{AO}} 2c_{lu}^* c_{mu}\left((ij|lm) - \frac{1}{2}(im|lj)\right) \tag{2.21}$$

using the short-hand notation

$$(ab|cd) = \int \phi_a^*(\mathbf{r}_1)\phi_b(\mathbf{r}_1)\frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|}\phi_c^*(\mathbf{r}_2)\phi_d(\mathbf{r}_2)d\mathbf{r}_1 d\mathbf{r}_2 \tag{2.22}$$

for the *two-electron integrals*. Semi-empirical methods make use of a series of approximations, which mainly reduce the effort to evaluate the two-electron integrals. This is done by expressing the two-electron integrals in form of parameters, which are specific for each chemical element. Furthermore, only valence shell electrons are taken into account, restricted to s- and p-electrons. Remaining electrons are summed up with the nuclear charge, yielding an effective charge for the nucleus. This affects the core Hamiltonian h_{ij} as well.

Semi-empirical methods either completely neglect the overlap of basis functions located at two different atoms (complete neglect of diatomic overlap, CNDO), or calculate only the overlap within the same molecular orbital (modified neglect of diatomic overlap, MNDO). At the same atom, there are 5 possible one-center two-electron integrals, which all have their specific parameter. These are used for the construction of all remaining two-electron integrals, together with a set of parameter derived from the form of the s- or p-orbital at the corresponding element.

The problem with the parametrization is the nuclear repulsion, which is not correct due to the effective nuclear charge. As a consequence, the nuclear repulsion needs to be parametrized, too. This is where the standard MNDO method and more advanced methods like PM6 [145] differ, as the latter employs a more sophisticated scheme than MNDO [11]. PM6, which has been used in this work, is a recent version of the methods AM1 [22] and PM3 [23], but contains no changes in the overall formalism.

2.1.3 Density functional methods

Hohenberg and Kohn have shown in 1964 [9] that it is not required to know the electronic wave function $\Psi_{el}(\underline{\mathbf{r}})$ for a given set of nuclear coordinates $\underline{\mathbf{R}}$ to obtain any systems' energy E. Instead, it suffices to know the electronic density ρ, which is

$$\rho(\mathbf{r}_1) = \int |\Psi_{el}(\underline{\mathbf{x}})|^2 d\omega_1 d\mathbf{x}_2 d\mathbf{x}_3 ... d\mathbf{x}_N := \rho(\mathbf{r}) \tag{2.23}$$

Once one knows ρ, it is theoretically possible to directly calculate E, as Hohenberg and Kohn have shown the system's energy to be a direct functional of ρ (1st Hohenberg-Kohn theorem),

$$E = E[\rho(\mathbf{r})] \quad , \tag{2.24}$$

with $\mathbf{r}$ being a simple three dimensional vector in space. Only the optimal ground state density will yield the minimum, ground state system energy. This is a variational principle for densities (for wave functions see (2.11)) (2nd Hohenberg-Kohn theorem)

$$E_{min} = E_0 = E[\rho_0(\mathbf{r})] \leq E[\tilde{\rho}(\mathbf{r})] \tag{2.25}$$

Therefore any (normalized) trial density $\tilde{\rho}$ will yield an energy which is higher or equal to the ground state energy. Theoretically, the electronic density is much easier to handle, as it only depends on $\mathbf{r}$, instead of the coordinates for each single electron.

As a first approach, we may consider the basic physical interactions, analogous to the WF ansatz (see (2.5)):

$$E_{DFT}[\rho] = T[\rho] + E_{ne}[\rho] + E_{ee}[\rho] \tag{2.26}$$

T is the kinetic energy, which will be discussed below in greater detail. E_{ne} is the attractive nuclei-electron interaction

$$E_{ne}[\rho] = -\sum_{A}^{N} \int \frac{Z_A \rho(\mathbf{r})}{|\mathbf{r}_A - \mathbf{r}|} d\mathbf{r} \tag{2.27}$$

while E_{ee} can be divided into a Coulomb expression J, an exchange expression K and electronic correlation. J is the electron-electron repulsion between the densities at $\mathbf{r}$ and $\mathbf{r}'$

$$J[\rho] = \frac{1}{2} \int \int \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r} d\mathbf{r}' \tag{2.28}$$

K and correlation are covered in the E_{xc} functional below.

The kinetic energy of the system in most practical realizations of DFT is given by a representation of the electrons in terms of orbitals, similar to the wave function approach. These Kohn-Sham-orbitals (ψ^{KS}) [26] work effectively as one-electron orbitals which are, analogous to HF, retrieved from the Kohn-Sham equations

$$\left[-\frac{1}{2}\nabla_1^2 - \sum_{A}^{N_A} \frac{Z_A}{|\mathbf{r} - \mathbf{R}_A|} + \int \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' + \frac{\delta E_{xc}[\rho(\mathbf{r})]}{\delta \rho(\mathbf{r})} \right] \psi_i^{KS}(\mathbf{r}) = \epsilon_i \psi_i(\mathbf{r}) \quad . \tag{2.29}$$

The density can then be derived from

$$\rho = \sum_{i=1}^{N} |\psi_i^{KS}|^2 \tag{2.30}$$

and the kinetic energy from

$$T[\rho] \approx T_S[\rho] = \sum_{i=1}^{N} \left\langle \psi_i^{KS} | -\frac{1}{2}\nabla^2 | \psi_i^{KS} \right\rangle \quad . \tag{2.31}$$

However, E_{xc} in (2.29) is unknown. This functional covers the error in the kinetic energy plus the remaining, non-classical electron-electron interaction. These are exchange (x) and correlation (c) energies. In the local density approximation (LDA) one writes

$$E_{xc}[\rho] = E_x[\rho] + E_c[\rho] = \int \rho(\mathbf{r}) \epsilon_x^{LDA}[\rho(\mathbf{r})] d\mathbf{r} + \int \rho(\mathbf{r}) \epsilon_c^{LDA}[\rho(\mathbf{r})] d\mathbf{r} \quad , \tag{2.32}$$

which is only exact for a homogeneous electron gas of constant density. As can be seen, the energies can also be displayed as energy per particle ϵ_x^{LDA} and ϵ_c^{LDA}.

The total energy expression then becomes

$$E_{DFT}[\rho] = T_S[\rho] + E_{ne}[\rho] + J[\rho] + E_{xc}[\rho] \tag{2.33}$$

E_{xc} is what current DFT research is mostly about, as it denotes the exchange-correlation energy, which separates the exact energy (2.26) from our working expression (2.33). In the rest of this section, the functionals used in this thesis will be presented.

BP86

Within the BP86 functional, the $E_{xc}[\rho]$ is split like in (2.32). The functional's name arises from the terms for $E_x[\rho]$ (B [27]) and $E_c[\rho]$ (P86 [28]). Both functional parts are based on (2.32). The LDA exchange part reads explicitly as

$$\epsilon_x^{LDA} = -\frac{3}{4}\left(\frac{3}{\pi}\right)^{1/3}\rho^{1/3} \quad . \tag{2.34}$$

LDA however is a crude approximation, as electron density does change within a molecule. LDA is thus improved by the Generalized Gradient Approximation (GGA), which accounts for the first derivative of the density. The B functional is a direct result of the GGA correction to LDA

$$\epsilon_x^B = \epsilon_x^{LDA} - \Delta\epsilon_x^B \tag{2.35}$$

with

$$\Delta\epsilon_x^B = \beta\rho^{1/3}\frac{x^2}{1+6\beta x\sinh^{-1}(x)} \text{ with } x = \frac{|\nabla\rho|}{\rho^{4/3}} \quad . \tag{2.36}$$

β is a factor fitted for rare gases.

A complete derivation of the correlation term $E_c[\rho]$ would exceed the limits of this thesis. At this point, it suffices to say that the P86 is based on the same basic approximations as B (LDA and GGA), and is also parametrized to fit to rare gas atoms, specifically Neon.

BP86 is computationally very efficient, and yields geometries comparable to other high accuracy methods (like B3LYP, see below). It was therefore used in this study for fast reoptimization of MD frames (*i.e.*, snapshots).

B3LYP

In contrast to the BP86 functional, B3LYP does not simply follow eq. (2.32) [29, 30]. Instead, the $E_{xc}[\rho]$ is of the form

$$E_{xc}^{B3LYP} = (1-a)E_x^{LSDA} + aE_x^{exact} + b\int \rho(\mathbf{r})\Delta\epsilon_x^B[\rho(\mathbf{r})]d\mathbf{r} + (1-c)E_c^{LSDA} + cE_c^{LYP} \quad (2.37)$$

LSDA stands for Local Spin Density Approximation and is equal to LDA for closed shell systems (which are our only concern in this thesis). E_x^{exact} denotes exact Hartree-Fock-like exchange energy, calculated using the KS orbitals. Finally, E_c^{LYP} is a rather accurate GGA correlation functional [31], fitted to the Helium atom and is of lengthy explicit form (thus omitted at this point). The parameters a, b and c are again fitted parameters, in particular, a = 0.2. B3LYP has become one of the most popular functionals, as it combines high accuracy for most interesting applications with relatively low computational cost. It is therefore used throughout this thesis for various purposes. Note that the E_c^{LYP} of the B3LYP functional implemented in the Gaussian package [40] is slightly different from the B3LYP used elsewhere.

BHandHLYP

BHandHLYP is a variant of B3LYP for which the parameters a, b and c are set to 0.5, 0.5 and 1, respectively. This carries more Hartree-Fock-like exchange than B3LYP and can thus be used for connecting DFT results with HF calculations for comparison.

CAM-B3LYP

The exchange potential V_x of B3LYP at long distances is not vanishing fast enough,

$$\frac{\delta E_x}{\delta\rho} = V_x \sim -\frac{a}{r} = -\frac{0.2}{r} \quad , \quad (2.38)$$

instead of the correct $-1/r$ behavior. The Coulomb attenuating method (CAM) is a way to correct the long-range behavior of B3LYP. The resulting functional is called CAM-B3LYP [32]. Here, the amount of exact HF exchange is gradually increased with r. CAM-B3LYP does not use the exchange and GGA correction seen in (2.37), but a sum of the exact HF exchange E_x^{exact} and B88 (2.35) instead. The respective weights of the two functionals are determined by two parameters α and β, which have to fulfill

$$\frac{1}{r} = \frac{1 - [\alpha + \beta \cdot \mathrm{erf}(\mu r)]}{r} + \frac{\alpha + \beta \cdot \mathrm{erf}(\mu r)}{r} \tag{2.39}$$

for the electron-electron interaction. erf is the error function and μ is a conversion parameter which adjusts the ratio $\alpha + \beta/\alpha$ with $1/r$. Yanai *et al.* have found that a set of $\alpha = 0.19$, $\beta = 0.46$ and $\mu = 0.33$ yields optimal results for their given test set [32]. The results have shown that CAM-B3LYP is optimal for long-range polarization effects, which for example occur in long chains.

2.1.4 Calculation of UV/vis spectra

For the electronic excitation spectra of the molecules presented in chapters 4 and 5, we require both *oscillator strengths* f_I and *vertical excitation energies* ω_I. There are various ways to obtain these values.

Displaying the oscillator strengths at their respective vertical excitation energy already forms a line spectrum. The creation of a broadened, more realistic spectrum is facilitated by applying certain types of broadening, which are presented in the corresponding sections later on (see eq. (4.1)).

Configuration Interaction

A wave function based approach to gain the UV/vis spectrum of a molecule is based on the configuration interaction (CI) method. CI allows for more than one Slater determinant (2.9) to participate in a molecular state, in other words, more than one electronic configuration is used in the trial function. The additional configurations are constructed by exciting one or more electrons from an occupied orbital a to a virtual orbital r. When only taking singly excited Slater determinants into account ("CIS" [33]), the trial function becomes

$$\Psi^{\mathrm{CIS}} = C_0 \Psi_0 + \sum_{a=1}^{N_{occ}} \sum_{r=N_{occ}+1}^{N_{virt}} C_a^r \Psi_a^r \tag{2.40}$$

containing all possible single excitations to the ground state virtual orbitals (N_{occ} and N_{virt} are the number of occupied and virtual orbitals, respectively). The coefficients C denote the participation of any determinant Ψ to the total wave function.

As the determinants are orthonormal, we can construct an eigenvalue problem for the coefficients,

$$\underline{\underline{H}}\ \underline{C}_I = E_I\ \underline{C}_I \tag{2.41}$$

with

$$H_{ij} = \left\langle \Psi_i | \hat{H}_{el} | \Psi_j \right\rangle \tag{2.42}$$

and $\underline{\underline{C}}$ being a $L \times L$ matrix with L being the number of singly excited determinants. Each column designates the set of coefficients for an excited state I. The excitation wavelength ω_I can be simply calculated from the energy difference to the ground state, the oscillator strength f_I is derived from

$$f_I = \frac{2}{3}\omega_I | \left\langle \Psi_0^{\mathrm{CIS}} | \underline{\hat{\mu}} | \Psi_I^{\mathrm{CIS}} \right\rangle |^2 \tag{2.43}$$

with the dipole operator

$$\underline{\hat{\mu}} = \sum_i \mathbf{r}_i \quad . \tag{2.44}$$

Time-dependent DFT

Most UV/vis spectra presented in this thesis are based on time-dependent density functional theory (TD-DFT). A detailed derivation of the TD-DFT formalism is given in [34]. We start from a time-dependent representation (compare (2.30)) of the density in form of the sum over all Kohn-Sham orbitals ($\psi_i = \psi_i^{KS}$):

$$\rho(\mathbf{r}, t) = \sum_{i,\sigma} f_i^\sigma |\psi_i^\sigma(\mathbf{r}, t)|^2 \tag{2.45}$$

with f_i^σ as occupation numbers of the corresponding orbital with σ indicating a specific electronic spin. The orbitals are obtained from solving the time-dependent Kohn-Sham equation (compare the time-independent form (2.29)),

$$\left\{ -\frac{1}{2}\nabla^2 + \nu(\mathbf{r}, t) + \int \frac{\rho(\mathbf{r}', t)}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' + \frac{\delta E_{xc}[\rho_\uparrow(t), \rho_\downarrow(t)]}{\delta \rho^\sigma(\mathbf{r}, t)} \right\} \psi_i^\sigma(\mathbf{r}, t) = i\frac{\partial}{\partial t}\psi_i^\sigma(\mathbf{r}, t) \quad , \tag{2.46}$$

with $\nu(\mathbf{r}, t)$ being the external potential. The exchange-correlation function is already in the adiabatic approximation, as the exchange-correlation functional $E_{xc}[\rho_\uparrow(t), \rho_\downarrow(t)]$ only depends on the space-dependent density (at a given time t).

The electronic perturbation is brought by a periodic $w(t)$, and the linear response of the density of a given spin is given by

$$\delta\rho^{\sigma}(\mathbf{r},t) = \sum_{i,j} \psi_i^{\sigma}(\mathbf{r})\delta P_{ij}^{\sigma}(\omega)\psi_j^{\sigma *}(\mathbf{r}) \quad . \tag{2.47}$$

Please note that here, ω corresponds to the frequency of the Fourier transformed perturbation $w(t) \rightarrow w(\omega)$. The response of the Kohn-Sham density matrix $\underline{\underline{P}}^{\sigma}$ is

$$\delta P_{ij}^{\sigma}(\omega) = \frac{f_j^{\sigma} - f_i^{\sigma}}{\omega - (\epsilon_i^{\sigma} - \epsilon_j^{\sigma})} \cdot \left[w_{ij}^{\sigma}(\omega) + \sum_{k,l,\tau} K_{ij,kl}^{\sigma,\tau} \delta P_{kl}^{\tau}(\omega) \right] \tag{2.48}$$

with the coupling matrix $\underline{\underline{K}}$. ϵ denotes the energy of a single KS orbital. In the coupling matrix, we find the changes of the self-consistent field (ν^{SCF}) when the charge density changes:

$$\begin{aligned} \delta K_{ij,kl}^{\sigma,\tau} = & \int\int \psi_i^{\sigma *}(\mathbf{r})\psi_j^{\sigma}(\mathbf{r}) \frac{1}{|\mathbf{r} - \mathbf{r}'|} \psi_k^{\tau}(\mathbf{r}')\psi_l^{\tau *}(\mathbf{r}') d\mathbf{r} d\mathbf{r}' \\ & + \int\int \psi_i^{\sigma *}(\mathbf{r})\psi_j^{\sigma}(\mathbf{r}) \frac{\delta^2 E_{xc}[\rho_{\uparrow}, \rho_{\downarrow}]}{\delta\rho^{\sigma}(\mathbf{r})\delta\rho^{\tau}(\mathbf{r}')} \psi_k^{\tau}(\mathbf{r}')\psi_l^{\tau *}(\mathbf{r}') d\mathbf{r} d\mathbf{r}' \end{aligned} \tag{2.49}$$

Since we are using the adiabatic approximation for E_{xc}, $\underline{\underline{K}}$ is time-independent. For the UV/vis spectra, the perturbation can be considered to be a time-dependent electric field $\mathbf{E}$, *e.g.*, along the z coordinate,

$$w(\mathbf{r},t) = zE_z(t) \quad . \tag{2.50}$$

To get the oscillator strengths and vertical excitation energies mentioned above, we require to compute the dynamic polarizabilities. These are given by (for the xz-component)

$$\alpha_{xz}(\omega) = -2 \sum_{i,j,\sigma}^{f_i^{\sigma} - f_j^{\sigma} > 0} x_{ji}^{\sigma} \mathrm{Re}(\delta P_{ij\sigma}(\omega))/E_z(\omega) \quad . \tag{2.51}$$

x_{ij}^{σ} are the elements of the $\underline{\underline{x}}$ matrix using the unperturbed Kohn-Sham orbitals

$$x_{ij}^{\sigma} = \left\langle \Psi_i^{\sigma} \left| \hat{x} \right| \Psi_j^{\sigma} \right\rangle \tag{2.52}$$

which can be equally constructed for y and z. With (2.48) for δP, eliminating the imaginary part, we can rewrite (2.51)

$$\alpha_{xz}(\omega) = 2\underline{\underline{x}}^* \underline{\underline{S}}^{-1/2} \{\underline{\underline{\Omega}} - \omega^2 \underline{\underline{1}}\}^{-1} \underline{\underline{S}}^{-1/2} \underline{\underline{z}} \quad . \tag{2.53}$$

The two new matrices

$$S^{\sigma,\tau}_{ij,kl} = \frac{\delta_{ik}\delta_{jl}\delta_{\sigma\tau}}{(f^\tau_k - f^\tau_j)(\epsilon^\tau_l - \epsilon^\tau_k)} > 0 \tag{2.54}$$

and

$$\Omega^{\sigma,\tau}_{ij,kl} = \delta_{ik}\delta_{jl}\delta_{\sigma\tau}(\epsilon^\tau_l - \epsilon^\tau_k)^2 + 2\sqrt{(f^\sigma_i - f^\sigma_j)(\epsilon^\sigma_j - \epsilon^\sigma_i)}K^{\sigma,\tau}_{ij,kl}\sqrt{(f^\tau_k - f^\tau_l)(\epsilon^\tau_l - \epsilon^\tau_k)} \tag{2.55}$$

are the expressions we need to evaluate the oscillator strengths f_I and vertical excitation energies ω_I. This is due to the sum-over-states expression for the mean dynamic polarizability

$$\bar{\alpha}(\omega) = \frac{1}{3}\mathrm{Tr}[\underline{\underline{\alpha}}(\omega)] = \sum_I \frac{f_I}{\omega_I^2 - \omega^2} \quad . \tag{2.56}$$

Using both (2.53) and (2.56), it is possible to derive ω_I by solving the eigenvalue equation

$$\underline{\underline{\Omega}}\ \underline{F}_I = \omega_I^2 \underline{F}_I \tag{2.57}$$

with $\underline{F}_I$ being the respective eigenvectors. Using these eigenvectors, the oscillator strengths can be calculated using

$$f_I = \frac{2}{3}(|\underline{x}^*\underline{\underline{S}}^{-1/2}\underline{F}_I|^2 + |\underline{y}^*\underline{\underline{S}}^{-1/2}\underline{F}_I|^2 + |\underline{z}^*\underline{\underline{S}}^{-1/2}\underline{F}_I|^2)/|\underline{F}_I|^2 \quad . \tag{2.58}$$

To characterize each excitation I, it is possible to obtain coefficients like in the CIS case using the vector $\underline{F}_I$ [35]. These can be used to analyze the excitations in terms of participating electrons, which will be done specifically in chapter 5.

2.1.5 Normal Mode Analysis and Thermal Correction

The formalisms given above yield electronic energies based on a fixed nuclear geometry, and together with the classical nuclear repulsion, a potential energy surface (PES) can be constructed. The masses of the nuclei, however, have an influence on chemical reactions: the molecule contains additional energy in form of zero point energy (ZPE). The value of the ZPE is dependent on the form of the PES and also on the masses of the nuclei, as shown below. ZPE is especially important for isotopic effects (chap. 3).

To account for the ZPE, we use the harmonic approximation for each of the $3N_A$ degrees of freedom with all the nuclei at the minimum of the electronic potential surface

$$V(\underline{q}) = \frac{1}{2}\sum_{i,j}^{3N_A} L_{ij} q_i q_j \tag{2.59}$$

with $\underline{q}$ being mass weighted displacement coordinates

$$q_i = \sqrt{m_i} \cdot \left(x_i - x_i^0\right) \tag{2.60}$$

x_i^0 being the equilibrium value of coordinate. $\underline{\underline{L}}$ is the Hesse matrix of the second derivatives of the PES along q_i. The off-diagonal elements of $\underline{\underline{L}}$ couple specific coordinate changes. By diagonalization of the Hesse matrix we obtain the decoupled eigenvectors, called "normal modes" $\underline{Q}$.

$$\frac{1}{2}\sum_{i,j}^{3N_A} L_{ij} q_i q_j \quad \rightarrow \quad \frac{1}{2}\sum_{i}^{3N_A} \lambda_i Q_i^2 \tag{2.61}$$

Normal modes represent orthogonal molecular movements.
Six of these, or five in linear molecules, represent rotational or translational movement with λ_i close to zero. All other λ_i are the squares of the harmonic vibrational frequencies ω_i for the normal mode Q_i. The ZPE can then be calculated from

$$E_{ZPE} = \frac{1}{2}\sum_{i} \omega_i \tag{2.62}$$

and for idealized problems at 0 K, this is directly added to the electronic energy.

For finite temperatures T, also higher vibrational states are occupied, and the molecule gains rotational and translational energy. Thus, the zero point energy cannot be used exclusively and we have to employ further corrections.

Any particle with the ability to thermally occupy more than one state ν will acquire a sum of states

$$q = \sum_{\nu}^{\text{all states}} e^{-E_\nu/(k_B T)} \tag{2.63}$$

which approaches 1 in the limit of 0 K, resulting in $E(T) = E_{el} + E_{ZPE}$ as stated above. For higher T, the sum increases and higher states are populated. Approximating the molecule as an ideal gas, the Gibbs free enthalpy of the system can be calculated from the inner energy U and entropy S at a given temperature T

$$G = H - TS = U + pV - U + A = pV - k_B T ln(q) \quad . \tag{2.64}$$

A is the Helmholtz energy, p the pressure and V the volume. The calculation of q_{sos} is omitted here, it should suffice to say that it consists of

$$q_{sos} = q_{tr} \cdot q_{vib} \cdot q_{rot} \cdot q_e \tag{2.65}$$

with q_{tr}, q_{vib}, q_{rot} and q_e being the contributions of translation, vibration, rotation and degeneracy of the electronic energy, respectively. The harmonic frequencies employed in the calculation of q_{vib} are scaled such that they are on average more consistent with the empirical values. The scaling factors depend on quantum chemical method and basis set. In this thesis, we used 0.983 for B3LYP/6-31G(d,p) [36] and 0.918 for HF/6-31G(d,p) [37].

2.1.6 Transition states

After the introduction of normal modes in the previous section, the concept of transition states will now be properly defined. A transition state (TS) is a maximum of a pathway on the potential energy surface connecting two minima. In an idealized case, we can assume the coordinate to be simply a single atom moving along a fixed coordinate, but in complex, biological systems, this is rarely the case. As such, the pathway will be the movement of a set of atoms along a set of coordinates, *e.g.* normal modes, see (2.61).

The $3N_A$-6(5) vibrational eigenvalues λ_i introduced in (2.61) are all positive for a minimum, as they are the second derivatives of the potential energy surface along their respective normal mode. This is not the case for a maximum, where the curvature is negative. More specifically, a TS of first order has a single negative eigenvalue λ_{TS} and thus an imaginary frequency ω_{TS}.

Eyring TS theory treats this specific mode classically. The other modes are treated quantum mechanically as normal modes as stated in the previous section 2.1.5. We can thus equally calculate G values for the minima and for the TS, too. We can then derive the activation free enthalpy

$$\Delta G_A = G_{\mathrm{ts}} - G_{\mathrm{min}} \tag{2.66}$$

and the rate constant k at a given temperature T for the specific reaction from a minimum passing the TS

$$k = \frac{k_B T}{h} \cdot e^{-\Delta G_A/(k_B T)} \tag{2.67}$$

The computational effort for finding a transition state can be similar to that of a simple geometry optimization, following the gradient along the desired transition mode upwards, not downwards. However, to find the set of atoms which participate in the movement, a normal mode analysis needs to be performed at least once at the beginning of the calculation, and should be regularly updated due to changes of coupling during the optimization. This method is for example implemented in the TURBOMOLE quantum chemistry package.

Another approach, implemented in the GAUSSIAN package, connects both minima with a parabola for a first guess of the transition state structure, and starts the search from the center of this curve [38]. In any case, the resulting transition state structure should be cross-checked by forward and backward propagation along the reaction coordinate, as these calculations must finish in the correct reactant/product states, if the transition state was correct.

In the simple cases presented in this thesis, however, a simple relaxed surface scan along a diatomic distance was sufficient. This is done by simply fixing the desired coordinate to a desired value and optimizing all other degrees of freedom. By stepwise changing the value of the fixed coordinate, a maximum along this respective coordinate can be found without the need of the sophisticated schemes stated above.

While we have now covered some improvements to the energetic analysis, we have yet to discuss possible improvements to the model itself. For biological molecules, the first improvement should be to include more of the immediate environment, like the solvent. The first step towards this are self consistent reaction fields, like the polarizable continuum model.

2.1.7 Polarizable Continuum Model

The first level of incorporation of the environment is the embedding of the solute (the QM system) into an approximated version of the solvent. Such approaches are called continuum models, and there are many varieties of these continuum models available.

All continuum models have in common that a cavity has to be constructed, in which the QM system will be located. This cavity is surrounded by a continuous dielectric medium to represent the solvent. However, it is clear that by this approach, we will miss some important effects; such as hydrogen bonds or other explicit electronic/molecular interaction. These can be incorporated with other means which will be discussed later.

The interactions must be calculated in a self-consistent scheme. This is due to the changes in dipole moment and polarization induced by both the medium and the molecule, which change for each calculation cycle. The methods are therefore called self-consistent reaction field (SCRF) methods. Here, we exclusively used the polarizable continuum model (PCM) [39] type of SCRF, which is implemented in the Gaussian09 package [40]. More specifically, this is the integral equation formalism variant (IEFPCM) [41], which is the default SCRF method in the package.

All PCM methods, including IEFPCM, build a "realistic" cavity around the molecule, as the cavity is based on a surface scan. The probe radius is a parameter defined specifically for the requested solvent. The resulting cavity is realistic in respect of having gaps at places where the probe is too large to fit in. However, such cavities are rarely spherical, which leads to a sophisticated description of the interaction between the molecule and the cavity surface. As this would exceed the limits of this thesis, we will stick to a few words regarding the general approach.

To gain the influence of the solvent within the IEFPCM formalism, we need to introduce a new potential term to the QM Hamiltonian, the potential $V_s(\mathbf{r})$, which arises from the electric field perpendicular to the cavity. This will be simply added to the system Hamiltonian

$$\hat{H} = \hat{H}_0 + V_s(\mathbf{r}) \tag{2.68}$$

The potential can be split

$$V_s(\mathbf{r}) = V_M(\mathbf{r}) + V_R(\mathbf{r}) \tag{2.69}$$

into a sum of both the electrostatic potential $V_M(\mathbf{r})$ produced by ρ_M, and $V_R(\mathbf{r})$ which is the reaction potential produced by the polarization of the medium. For the IEFPCM, the potentials are defined with help of proper Green functions $G(\underline{x}, \underline{y})$. These functions

denote the potential at position $\underline{x}$ produced by a charge at $\underline{y}$.

The problem is the definition of $V_R(\mathbf{r})$. PCM methods are a group of apparent surface charge methods (ASC). In ASC methods, an apparent surface charge density $\sigma(\mathbf{s})$ occurs at a position $\mathbf{s}$ along the cavity surface Γ:

$$V_\sigma(\mathbf{r}) = \int_\Gamma \frac{\sigma(\mathbf{s})}{|\mathbf{r}-\mathbf{s}|} d\mathbf{s} \tag{2.70}$$

which is an exact form for $V_R(\mathbf{r})$[2]. The integral is however computationally demanding, so that the surface is split into several "tesserae" in which $\sigma(\mathbf{s})$ is considered a constant. Logically, the tesserae have to be small enough to avoid the generation of a too large error. For a tessera k, we can multiply the charge σ_k with the tessera area A_k to gain a point charge q_i. This way, we change the description of the PCM from a continuous surface to a simplified set of point charges. The potential generated by these point charges can be simply summed up:

$$V_\sigma(\mathbf{r}) \approx \sum_i \frac{q_i}{|\mathbf{r}-\mathbf{s}_i|} \tag{2.71}$$

With having a small overview on the generation of a PCM, the question arises if the concept of point charges can be used for other applications as well. Instead of generating approximate point charges for a continuous medium at a molecule surface, we can think of generating the point charges from other sources, too. This is called electronic embedding and will be discussed now.

2.1.8 Electronic embedding

The point charges presented in the last section can also be generated by other methods. Most prominent in this respect are force fields (see sec. 2.2), where the point charges are directly available from the atom parametrization. A way to generate accurate point charges from quantum chemical systems is electronic population analysis using the electrostatic potential scheme, ESP [42, 43].

The implementation of the point charges in the Hamiltonian [44, 45, 46] is analogous to the one presented in the PCM section 2.1.7, (2.71). The electronic Hamiltonian gets

[2]Formally, $\mathbf{s}$ is of course a surface coordinate which needs some further attention to be compatible with the spatial coordinate $\mathbf{r}$.

extended by a potential term including all the point charges and their distance to each electron in the system. The classical Coulomb correction for the nuclear energy has to include the point charges, too.

Most of the long-range interactions between molecules are covered by including the Coulomb interaction. The electronic embedding scheme is however not the method of choice, as it can produce severe artifacts if the border area is not well defined. For example, acid/base pairs may create geometric distortions when optimized within a static electronic embedding scheme. This is due to the fact that the van-der-Waals interactions are omitted, which are the repulsive component for a proton transfer. In the worst case, protons or other atoms may just sit on top of a point charge at the end of a geometry optimization.

In the next part of this chapter, we will therefore present the classical approach for describing long range molecular interaction, namely force fields. Here, both Coulomb and van-der-Waals forces are included, yet we will lose the exact electronic properties of our model systems.

2.2 Force Field Theory and Application

Force fields describe the properties of a system exclusively on the basis of empirical potentials. They make use of parameters which have been defined beforehand [11, 12, 13, 14]. These parameters will usually not change during the calculation. Parameters are *e.g.* typical relaxed bond angles, harmonic force constants and so on. *I.e.*, using an harmonic potential for atomic bonds means that no covalent bonds can be formed or broken. As this is the case in this thesis, all atoms retain their chemical character and bond partners in the calculations presented here.

While also being strongly limited in terms of accuracy, force fields are in the end the method of choice for any system larger than a thousand atoms. Recent developments however seek to incorporate more advanced molecular interactions, with the help of polarizable force fields [47]. Force field based calculations are generally called MM (molecular mechanics) methods, in contrast to QM. This section will discuss the force fields and their energy expressions used in this work. We will thereafter present the basics required to

understand the molecular dynamics simulations using force fields (MD, sec. 2.2.2).

2.2.1 Force Fields

To calculate the energy of a molecular system classically, one has to rely on previously fitted *parameters*. These are either defined for for pairs of atoms (covalent bond, Coulomb and van-der-Waals interaction) or small sets of atoms (covalent bond angles, dihedrals or out-of-plane vibrations). A corresponding potential energy expression

$$V_{\mathrm{FF}} = V_{\mathrm{bonds}} + V_{\mathrm{angles}} + V_{\mathrm{dihedrals}} + V_{\text{non-bonding}} + V_{\mathrm{other}} \tag{2.72}$$

and the respective parameters for each of the terms in (2.72) are called a *force field*.

In this thesis, we used both the Charmm27 [12] and Amber99 [13] force fields, which were employed in the Tinker [48] and Gromacs [49] packages.

Charmm

*C*hemistry at *Har*vard *m*olecular *m*echanics (Charmm) force fields have been developed for several applications [12]. The version used in this thesis, Charmm27 [50], was specifically designed for biological molecules such as DNA, RNA, lipids and peptides/proteins.

The Charmm27 potential is of the form

$$\begin{aligned} V_{FF}(\underline{\mathbf{R}}) = & \sum_{\text{bonds}} k_b(b-b_0)^2 + \sum_{\text{angles}} k_\theta(\theta-\theta_0)^2 \\ & + \sum_{\text{dihedrals}} k_\chi(1+cos(n_\chi\cdot\chi)) \\ & + \sum_{\text{Urey-Bradley}} K_{UB}(S-S_0)^2 \\ & + \sum_{\text{impropers}} k_\phi(\phi-\phi_0)^2 \\ & + \sum_{\text{non-bonded atom pairs}} \left(4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \frac{q_i q_j}{\epsilon_D r_{ij}}\right) \quad . \end{aligned} \tag{2.73}$$

The term is a sum of bond, bond angle, dihedral angle, Urey-Bradley, improper dihedral, van-der-Waals and Coulomb potentials, with k denoting the respective force constants. b is the bond length, θ the bond angle, χ the dihedral angle, S the distance between

two atoms which are two bonds apart (Urey-Bradley 1,3-distance), ϕ the improper dihedral angle and r_{ij} the non-bonded atom distance. Where appropriate, the index 0 denotes the equilibrium distance or angle. In the dihedral potential, n is the multiplicity. The van-der-Waals interaction is represented as Lennard-Jones (LJ) potential with σ_{ij} as the point of minimal interaction between i and j and ϵ_{ij} as the LJ well depth. For a pair of atoms ij, ϵ_{ij} is determined by taking the geometric mean of the parameters for each atom ($\epsilon_{ij} = \sqrt{\epsilon_i \epsilon_j}$), and the arithmetic mean for σ_{ij} ($\sigma_{ij} = (\sigma_i + \sigma_j)/2$). Finally, the Coulomb potential depends on the partial atomic charges q_i and the dielectric constant ϵ_D (which is 1 for all our concerns since we treat the solvent explicitly).

While most of the terms are very straight forward, the improper dihedrals and Urey-Bradley terms are included for a better fit to experimental vibrational spectra. Improper dihedrals denote out-of-plane vibrations of aromatic rings. Urey-Bradley terms refine the 1,3 repulsion occurring within a bond angle potential.

The Charmm27 force field was used within the Tinker molecular dynamics package [48].

Amber

Amber stands for *a*ssisted *m*odel *b*uilding with *e*nergy *r*efinement. Similar to most other force fields, Amber is both the name of a set of force fields and a molecular dynamics package [13]. We will refer to the force field Amber99 only [51]. Just as the Charmm27 force field, Amber is specifically aimed at predicting structural behavior of biological systems.

The force field term is identical to that of Charmm27 (2.73), apart from the Urey-Bradley and improper terms, which are omitted. In the official sources [52], the equation is denoted somewhat differently, but the alterations needed to convert the two force field terms into each other are trivial. One must be aware that this conversion affects parameters as well, resulting in the need of special "porting" of the parameter files. The Gromacs molecular dynamics package [49], which was used in this work, relies for example on a different format for the parameters than the standard Amber definition (*e.g.*, Lennard-Jones potential instead of a repulsive term and an attractive London force term). To transport the Amber99 parameters into the Gromacs package, we used an available Amber99 port [53].

Having a clear definition of the potential, finding energetic minima is now just a problem of choosing an appropriate numerical approach, such as steepest descent. It is however not trivial to do the dynamics at ambient temperature, which will be discussed in the next section. For the sake of simplicity, the periodic boundary conditions (PBCs) which are technically required for the minimization, too, will be discussed in the next section as well.

2.2.2 Molecular Dynamics

Molecular dynamics (MD) usually refers to the evolution of a molecular system in time by means of a force field description. MD simulations are very useful in the respect of biological applications, as biological systems in general are too large to be described by (even approximated) quantum dynamics. This purely classical approach comes with all the disadvantages of the force field description (see sec. 2.2), but it yields information which is not available otherwise. Most prominently, MD simulations give an overview on the flexibility of protein regions, amino acids and single hydrogen bonds. A statistical analysis over hundreds or thousands of "frames", recorded time steps, can show changes in this behavior depending on starting conformation.
Furthermore, MD simulations can expand the structural basis on which quantum chemical calculations are performed. The acquisition of the MD frames however is connected with a series of problems. The force field related problems have been discussed before in sec. 2.2. The problems arising from the MD setup are as follows:

- Periodic boundary conditions (PBCs) - For a realistic system, identical images of the simulation box are seen by each atom of the system. Thus, artifacts arising from the box edges are prevented. The implementation is very straight forward and will not be discussed further.

- Finite time step - The distance each atom moves from one calculation step to the other is defined by the time step times the atom's current kinetic vector (velocity). It is thus not continuously updated, and movement of atoms is not continuous as well. It needs to be continuous *enough* however, so that the atom may for example not unrealistically penetrate potential barriers just by appearing on the other side of the barrier. The time step must therefore be appropriately small.

- Initial conditions - In this thesis, all systems were instantly taken to 300 K from their potential minimum (corresponding to 0 K). This was done by generating random vectors at random directions corresponding to a velocity distribution expected at 300 K. The basis for this was a random seed, an integer.

- Choice of propagator - In order to achieve energy conservation within the system, an appropriate propagator needs to be chosen (see below).

- Temperature and pressure coupling - In this thesis, exclusively temperature coupling was used, and pressure coupling will be neglected. This setup is called "NVT", indicating the number of particles (N), the volume (V) and the temperature (T) to be constant. It might however be desired to have slight fluctuations in T to have more realistic dynamics. Thus, in order to keep the system at an approximately constant temperature, the system may take up heat from or dissipate heat to an external bath.

- Infinite long-range interactions - Interactions between non-bonded atoms must be approximated starting at a certain distance, as otherwise, the system would have an infinite number of terms to calculate when using PBCs.

As stated above, periodic boundary conditions are easy to implement, as they require a simple translation of the system along the box vectors and only affect the non-bonding interactions. For the time step, a convergence test regarding the investigated feature can determine an appropriate time step. The initial conditions have to be varied and checked for convergence on the property of interest. A stepwise heating of the system should also be done to see that the initial conditions do not create artifacts similar to those using a too small step size.
Three problems of MD simulations will be discussed in greater detail in the rest of this section.

Choice of Propagator

The propagator used for the MD simulations in this work is either the velocity verlet (Tinker [48]) or the leap-frog integrator (Gromacs [49]).

The velocity verlet integrator [54] is a symplectic operator, able to conserve the energy of a system in time. For the equations of motions to be solved, we need to update our coordinate vector $\underline{q}(t)$ with elements $q_i(t)$ to a new vector $\underline{q}(t+\Delta t)$. At the same time, we need to update the velocities along each coordinate to $\underline{v}(t+\Delta t)$ and the forces to $\underline{F}(t+\Delta t)$.

First, we calculate the force at time t

$$F_i(t) = -\frac{\partial V_{FF}(\underline{q}(t))}{\partial q_i} \tag{2.74}$$

using the force field potential term V_{FF}. Assume the time step to be small enough, the force $F_i(t)$ can be considered constant until reaching $q_i(t+\Delta t)$. We can thus calculate the new coordinates by

$$q_i(t+\Delta t) = q_i(t) + v_i(t)\Delta t + \frac{F_i(t)}{2m_i}(\Delta t)^2 \tag{2.75}$$

with m_i being the mass of the atom. Then, the force is updated at the new positions:

$$F_i(t+\Delta t) = -\frac{\partial V_{FF}(\underline{q}(t+\Delta t))}{\partial q_i} \tag{2.76}$$

To get the new velocities, we average the forces at $q_i(t)$ and $q_i(t+\Delta t)$, to gain a better approximation for the real updated impulse. The new velocities are then

$$v_i(t+\Delta t) = -\frac{F_i(t)+F_i(t+\Delta t)}{2m_i}\Delta t + v(t) \quad . \tag{2.77}$$

The leap-frog integrator [55] used by Gromacs is equivalent to the verlet integrator [56] (not to be confused with the *velocity* verlet integrator already presented). We first need to evaluate the forces at time t, just as in (2.74). We then update the velocities not at time $t+\Delta t$, but at $t+\frac{\Delta t}{2}$:

$$v_i\left(t+\frac{\Delta t}{2}\right) = v_i\left(t-\frac{\Delta t}{2}\right) + \frac{F_i(t)}{m_i}\Delta t \tag{2.78}$$

Since we update v_i at $t+\frac{\Delta t}{2}$, we need to use $v_i(t-\frac{\Delta t}{2})$ in (2.78). We then update the positions at $t+\Delta t$ by

$$q_i(t+\Delta t) = q_i(t) + v_i\left(t+\frac{\Delta t}{2}\right)\Delta t \quad . \tag{2.79}$$

The name "leap-frog" arises from the fact that positions and velocities are not updated at the same time, but shifted by $\frac{\Delta t}{2}$. The update cycle of one property "leaps" over the update points of the other property.

Temperature Coupling

Temperature coupling is performed by a so-called thermostat. In the Tinker suite, we relied on the Berendsen thermostat [57], while in the Gromacs case, v-rescale [58] was used. Temperature coupling suppresses the changes of kinetic energy in the ensemble, which necessarily occur during the propagation with the integrators given above.

In the Berendsen thermostat [57], the temperatures are scaled such that the change in T is

$$\frac{dT}{dt} = \frac{T_0 - T}{\tau} \tag{2.80}$$

with T_0 as the desired temperature of the ensemble and τ a user-specified time constant. The time constant is usually in the range of several fs to 1 ps, and was set to 0.1 ps for all purposes in this thesis. From τ, we can derive a scaling factor λ

$$\lambda = \left[1 + \frac{\Delta t}{\tau_T}\left(\frac{T_0}{T_{\text{old}}} - 1\right)\right]^{1/2} \tag{2.81}$$

which is multiplied to the updated velocities of (2.77) or (2.78). T_{old} is taken from the kinetic distribution of the old velocities. The connection between τ and τ_T is

$$\tau = \frac{2C_V\tau_T}{N_f k_B} \tag{2.82}$$

with C_V being the heat capacity of the system, N_f the total degrees of freedom of the system and k_B Boltzmann's constant.

The Berendsen thermostat will however not generate a correct *distribution* of the kinetic energy. The v-rescaling thermostat [58] extends the Berendsen thermostat by introducing a correction term for the kinetic energy change, which ensures the correct kinetic energy distribution.

Long-range Interactions

Long range interactions in a finite box pose a fast scaling problem (N^2). With periodic boundary conditions, the problem is even worse, since technically, an infinite number of pair potential terms would have to be cared for. Sophisticated methods are available to treat these problems exactly, with scaling as low as $N \cdot ln(N)$ (particle mesh Ewald, PME [59]). This is the method used throughout this thesis for the long range electrostatics.

Short-range electrostatics within a (chosen) distance of 1.2 nm were treated explicitly. Analogously, van-der-Waals forces were evaluated using the a simple cut-off scheme, neglecting terms with a distance larger than 1.2 nm between two atoms.

We have up to now shown the basic techniques for the evaluation of both quantum mechanical and classical energies. We will now take a brief look at the methodology used in this thesis for the combination of both in form of a QM/MM approach.

2.3 QM/MM methods

QM/MM is the coupling of both QM and MM methodologies for energy calculations. There are two general ways to accomplish the coupling, one being the *additive*, the other being the *subtractive* approach. Both approaches are based on a layering of the system: We define one or more core regions, which are treated with computationally demanding quantum chemical methods. The remainder of the system will be described by one or more less expensive methods, such as semi-empirical QM or force fields. While the treatment of the single layers is trivial, the critical question is how the borders and the interaction between these layers are computed. This is where the two approaches differ.

2.3.1 Additive Approach

The additive approach is using a coupling term $E_{\text{QM-MM}}$ to evaluate the interactions between the QM and MM regions for the total energy of the system:

$$E_{\text{tot}} = E_{\text{QM}} + E_{\text{MM}} + E_{\text{QM-MM}} \tag{2.83}$$

Here, E_{QM} and E_{MM} represent the energies of the QM and MM regions, respectively, treated with the corresponding methods. In this thesis, the additive approach was not used, so we will not discuss the construction of the $E_{\text{QM-MM}}$ in detail. Yet, it should be stated that the coupling term poses much less of a problem when there are no covalent bonds connecting the layers. Having no bonds to cut leaves us with the van-der-Waals and Coulomb interactions only, of which the latter are easy to implement. If bonds need to be cut, the bond is saturated by a link atom (mostly hydrogen) in many calculations. The treatment of cut bonds at borders is a field of ongoing research. An extensive review

centered on the additive approach can be found in ref. [15].

2.3.2 Subtractive Approach

For our purposes, we used the computationally slightly more demanding *subtractive approach*. One of the most prominent representatives of the subtractive approach is the ONIOM scheme [60], implemented in Gaussian [40]. We however used the ComQum [61] program, since we required to couple both the Amber [13] and TURBOMOLE [62] programs.
Subtractive approaches couple the layers by calculating the full system energy (the "real" system) and the QM region (the "model" system) energy with the MM method. This way, we can subtract the energies, yielding the total system energy

$$E_{\text{tot}} = E_{\text{model, QM}} + E_{\text{real, MM}} - E_{\text{model, MM}} \quad . \tag{2.84}$$

The most obvious difference to the additive approach is that the interactions are present on the MM level only. It can be seen that the choice of a MM method for the latter two terms in (2.84) is purely arbitrary, allowing also for, *e.g.*, a QM/QM scheme. One could use a more demanding DFT method like B3LYP for the QM part and a less demanding DFT method like BP86 for the other calculations.
As in the additive approach, cut bonds at borders are generally saturated with hydrogens. The problem with the cut bonds however remains, if the energetic contributions from the MM method for a link atom are not approximately the same as for the QM method. If that is the case, the contributions from the two levels do not cancel out in the difference, and energetic differences of two geometries (*e.g.* activation energies) may be wrong.

A summary of the approaches presented in this chapter can be seen in figure 2.1. Concluding this chapter, we have to state that none of the methods represents the best of the best for every application. We will therefore state explicitly which approaches were used for each of the systems, which will be presented in the next chapters.

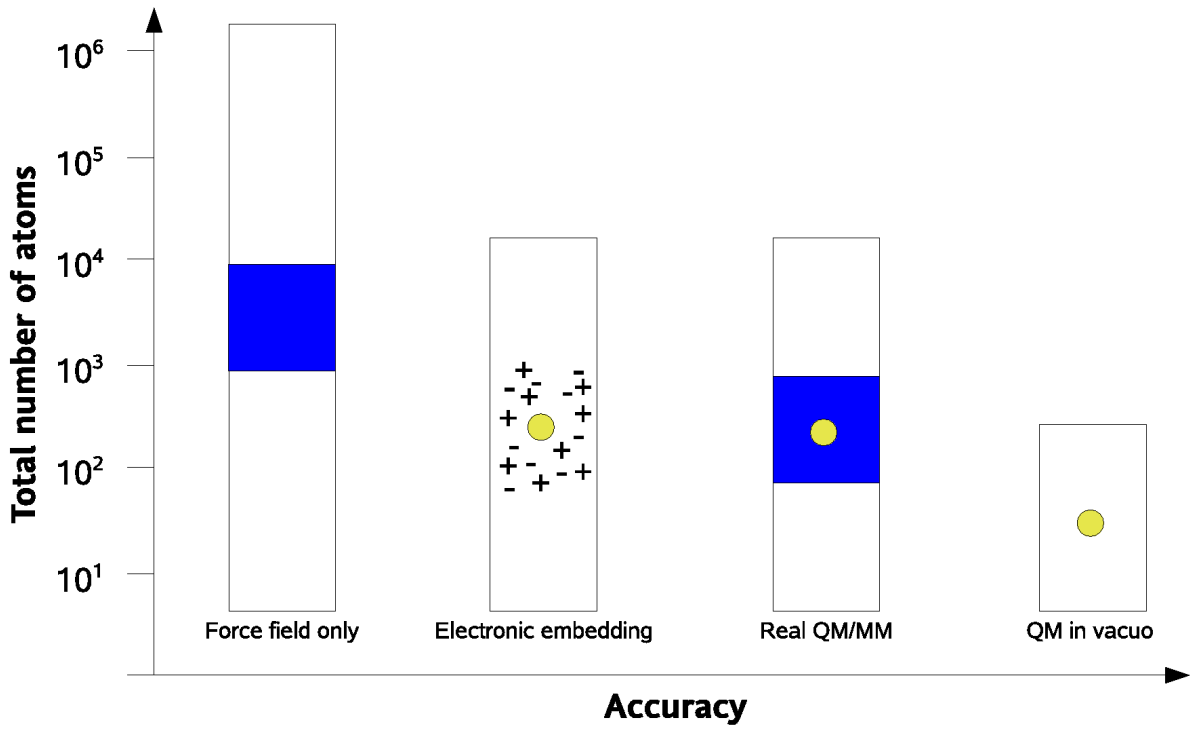

Figure 2.1: Schematic depiction of all levels of model complexity used in this thesis. The large surrounding boxes indicate the range of system sizes for which the approach can be applied. A sketch of the "zones" present in each model is given for each box. Yellow: QM region, Blue: MM region, +/-: Coulomb point charges. In all cases, the real upper limit depends on available computational resources.

Chapter 3

The RuBisCO Enzyme

3.1 The RuBisCO carbon isotope effect

The overall reaction of Ribulose-1,5-bisphosphate Carboxylase/Oxygenase (RuBisCO, EC 4.1.1.39[1], [63], fig. 3.1) exhibits a kinetic isotope effect regarding the reactivity of the CO_2 carbon atom [18] with the total reaction rates k_{12} and k_{13} depending on mass

$$\alpha = \frac{k_{12}}{k_{13}} \tag{3.1}$$

which is 1.030 ± 0.001 for spinach ($\hat{=}$ 3‰; at room temperature, pH=8 [18]) and of similar order for other organisms [64]. Unfortunately, apart from the diffusion processes in and out of the active site, RuBisCO features a series of reaction steps. Yet, the carboxylation step itself, in which the CO_2 carbon directly forms a bond [65], is supposed to be most important for the isotope effect.

In this chapter, the overall properties of RuBisCO and some related research fields will be presented. Thereafter, the active site of RuBisCO will be discussed to construct appropriate model systems for the carboxylation step. The different models will then be presented, as well as the results for the isotopic effect obtained from the respective calculations. A short overview will finish this chapter.

3.1.1 RuBisCO: Function and Abundance

RuBisCO is a soluble lyase, and in terms of molecular weight the most abundant protein on the planet [66, 67]. RuBisCO is responsible for the carboxylation of ribulose-1,5-

[1] *Enzyme Commission number*, describing the reaction catalyzed by the enzyme. RuBisCO is a lyase (4), acting on a carbon-carbon bond (1) forming a carboxyl group (1). The last number is a running index.

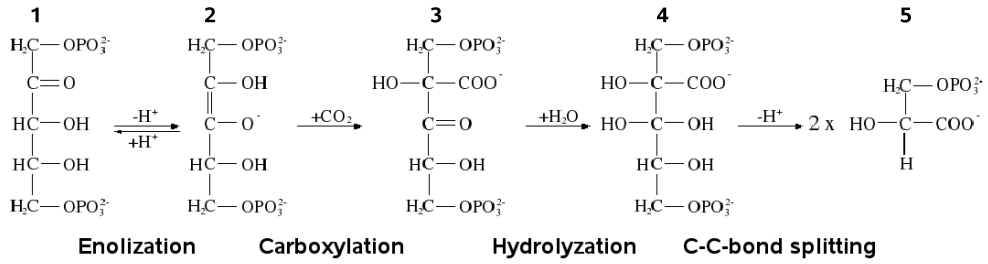

Figure 3.1: Carboxylation reaction pathway of RuBisCO according to Refs.[70, 65], adapted from [71]. Individual steps: Enolization, followed by the carboxylation step and fast hydrolyzation, and finally C-C bond cleavage [18]. Compounds numbered for reference in the text.

bisphosphate (RuBP, 1 in fig. 3.1): This is performed by an electrophilic attack of carbon dioxide to an enolized form of RuBP (2). Following this reaction, a hydroxyl group from a water molecule is transferred to the carboxylated compound (3). Finally, after the formation of a gem-diol (4), the weakened bond between the C2 and C3 carbon of the intermediate is broken (for atom labels, see fig. 3.3). One of the two cleavage products is a carbanion, which is stereo-specifically protonated. Both reaction products are finally released as 3-phosphoglycerate (3PGA, 5 in fig. 3.1). More details on the reaction can be found in various references [65, 68], since it is one of the central reactions which sustains life on earth and is therefore intensively studied. This is the reaction which is termed the *carbon fixation step of the Calvin-Benson cycle* [69]; it takes place in photosynthetic bacteria and plant/algae chloroplasts. A scheme of the carboxylation reaction is found in fig. 3.1 as well as a schematic overview of the Calvin-Benson cycle in fig. 3.2. For a more detailed picture, refer to fig. 1.1 in sec. 1.2.

For practical purposes, RuBisCO seems to be the kinetic bottleneck of carbon fixation [64]. It not only has a turnover rate of only 1.75 s^{-1} (for *Spinacea oleracea* per active site [72]), it also exhibits a lot of side reactions, resulting in unproductive pathways and/or self-inhibition processes [73, 74]. Improvement of RuBisCO in any way is therefore a target for many scientists all over the world.

3.1.2 Active Site of RuBisCO

We will use in the following amino acid numbering as in the RuBisCO of *Chlamydomonas reinhardtii* (*C.r.*; PDB dataset 1GK8 [75]), depicted in fig. 3.3. The RuBisCO active

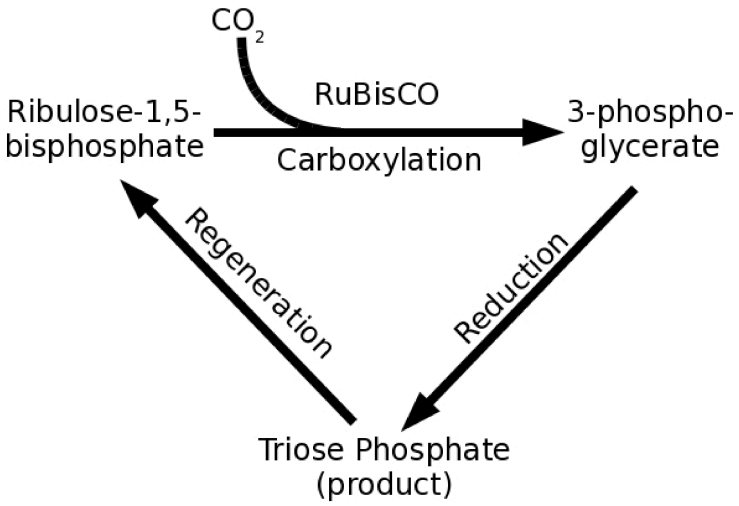

Figure 3.2: Calvin-Benson cycle overview, reduced to the three key processes. Note that the reduction and regeneration phases are multiple reactions catalyzed by various enzymes. Neither export processes nor NADPH or ATP consumption are shown.

site is composed of a magnesium ion (2+ charge), which has a coordinated complex of three sequentially close acidic amino acids (KCX201, Asp203 and Glu204). The first of these amino acids is a modified lysine. It has been carboxylated with CO_2 resulting in an acidic carboxyamino group suitable for forming the Mg^{2+} complex [65]. While the remaining three of six coordination sites can be saturated by water, during the reaction two of these sites are required for binding of the O2 and O3 of RuBP (see fig. 3.3B). The last remaining coordination site is assumed to be used for the binding of carbon dioxide. For details on the CO_2-binding process, see sec. 3.1.3.

Many other amino acids bind to the hydroxyl and phosphate groups of RuBP, which will not be discussed at this point. However, for disambiguation, it should be noted that the active site is composed of two large RuBisCO subunits. We will call the main subunit, which holds the magnesium, “subunit A” and the other, supporting subunit “subunit B” (see fig. 3.3 for comparison). It should be noted that a dimer of the two large subunits contains a C_2 axis, and thus it contains two active sites. For each site, one subunit is the A subunit and the other the B subunit, and *vice versa* for the other active site. This situation is depicted in fig. 3.4 for clarification.

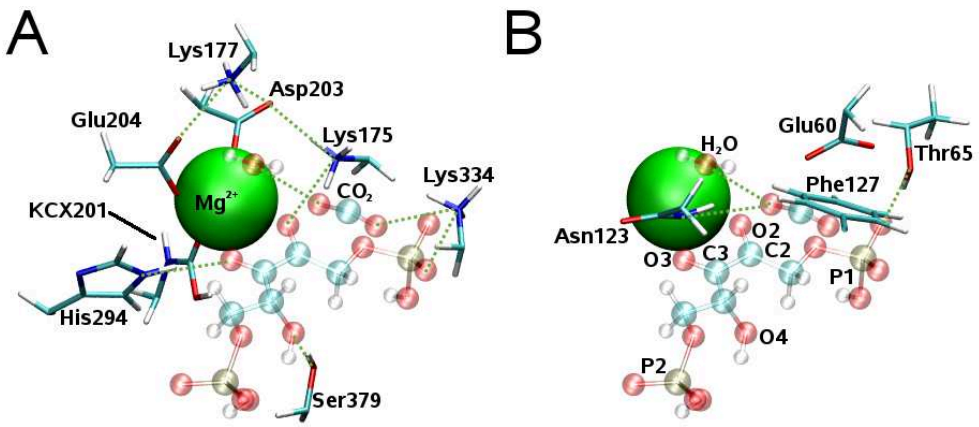

Figure 3.3: Active site of *Chlamydomonas reinhardtii* RuBisCO, based on PDB structure 1GK8 [75], in a 2L version kindly provided by Prof. Andersson, Uppsala, Sweden. Modified to fit the reaction site geometry proposed by Kannappan and Gready [76]. Note that the active site cannot be fully located on a single large subunit, and display of both stacked subunits is required. Only protein structures in contact with the ribityl-chain and magnesium ion are shown. **A**: Amino acids of main subunit A; **B**: Amino acids of supporting subunit B. Green dotted lines indicate hydrogen bonding, the hydrogen bond between Lys334 and Glu60 is not displayed. Non-amino acid labels represent reactant atoms or molecules.

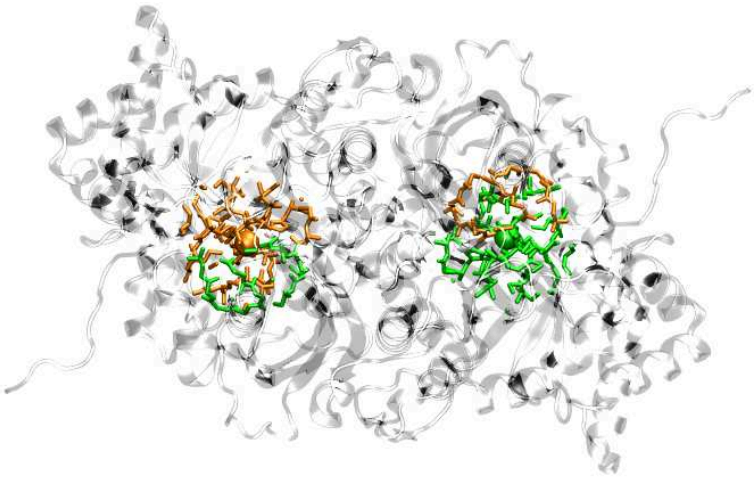

Figure 3.4: 2L sub-complex (2 large subunits) taken from the hexadecameric 8L8S *Chlamydomonas reinhardtii* RuBisCO complex, kindly provided by Prof. Inger Andersson, Uppsala, Sweden. 10 Å spherical zone around Mg^{2+} ion drawn as an approximation for the active site, colored according to corresponding subunit. The active sites clearly extend over both subunits.

3.1.3 Carboxylation Step and Model Construction

Carbon dioxide is essentially coordinated to the magnesium ion, to be close to the enolized RuBP C2-C3 double bond. In conventional models [65, 80], the carbon dioxide is coordinated directly to the ion, as it is indicated by crystal structures found by using carboxyarabinitol-1,5-bisphosphate [81] (CABP, a transition state analogue). This requires the replacement of a water molecule, which is assumed to occupy this coordination site before CO_2 binding.
While these CABP structures indicate CO_2 to be directly bound to the magnesium ion, recently an alternative model has been proposed stating that the exchange of water for the carbon dioxide on the binding site may be unnecessary [76]. Here, the water molecule coordinated to Mg^{2+} remains in place, and carbon dioxide itself is bound to the complex via the water molecule. This idea is intriguing, as it allows for closing of the site without the need of further water relocation processes. In addition, this model is able to explain the origin of the water molecule required for the hydrolysis reaction step.

Our model construction was thus based on two different active site geometries, and we compare both models in terms of reactivity and CO_2 carbon isotope effect.

The overall procedure for each of the mechanisms was as follows: A model of the carboxylation geometry was constructed (see below), and then the relaxed potential energy curve was followed along the C2(RuBP)-C(CO_2) distance coordinate. Explicit stepwise scans were performed (with a step size of 0.1 Å, then 0.01 Å for further refinement) to locate the stationary points along the reaction coordinate. The maxima and minima along this reaction pathway were characterized by normal mode analysis, and frequencies scaled according to sec. 2.1.5 for the calculation of the energies (using 298.15 K and 1.0 atm for the calculation of thermodynamical properties). The differences in the isotopic composition were included by using different isotopes (^{12}C or ^{13}C) for the CO_2 carbon.
As results from our calculations, we gained thermally corrected free electronic energies for the reactant (r), transition (ts) and product (p) state. We formed differences (according to (2.66)) for the acquisition of both the activation barrier

$$\Delta G^{\ddagger} = G_{ts} - G_r \tag{3.2}$$

and analogously for the free enthalpy of the reaction

$$\Delta G = G_p - G_r \tag{3.3}$$

From (3.2), with (2.67) we can calculate absolute reaction rates k. Finally, the isotopic effect α was calculated according to (3.1) using the obtained reaction rates for each isotope.

The procedure was performed for both HF and B3LYP levels of theory using the 6-31G(d,p) basis set. The latter basis was also used for the calculation of the systems within a polarized continuum model (PCM), using chloroform ($\epsilon = 4.7113$) as solvent. This is reported to approximately mimic the dielectric properties of a protein environment [82]. The resulting energy differences and barriers are presented in secs. 3.2.1 and 3.2.2.

3.2 Models of the Carboxylation Step

3.2.1 Conventional Model

Model Construction

We constructed the "conventional model" based on refs. [65, 80] using most amino acids of subunit A, which are in direct contact to the reactants. We restricted our model to the functional residues, omitting all backbone and aliphatic chain atoms. Explicitly included structures were RuBP and CO_2, Mg^{2+}, Lys175, Lys201 (carboxylated, "KCX201"), Asp 203, Glu204, His294, Arg295, His327, and Lys334. Coordinates were taken from PDB structure file 1IR1 (spinach, [81]), which contained the highest resolution RuBisCO structure at the time of the model construction. RuBisCO active sites are highly conserved, thus the source of geometry data only plays a minor role. The hydrogens were added to have all acid and base amino acids charged, including the His294 residue. The RuBP phosphate groups were singly protonated, at the oxygen having the longest P-O bond length in the crystal structure. To account for structural waters, the water molecules labeled 31, 91, 101, 141, 347 and 541 in the PDB dataset were included. This model consisted of 102 atoms and is shown in fig. 3.5, denoted as model "1" in what follows.

A smaller model was constructed using only waters to fill the coordinate sphere of the

Mg ion instead of the amino acids, and only including O2, O3, C2, C3 and the carbon dioxide, filling missing valences with hydrogens. This model consisted of 19 atoms and is shown in fig. 3.6, denoted as model "1b" in the following.

System treatment

For all calculations of the 102 atom system, we had to freeze all heavy atoms except for the atoms of CO_2, C2, O2, O3 and O4 of RuBP and all hydrogens. This is due to unphysical geometry distortions of the amino acids arising from the missing environment. The choice of unfrozen atoms was based on retaining the overall geometry, while giving the reacting atoms enough flexibility to form the required bond. These restrictions were not needed for the 19 atom system.

Results

It was not possible to produce a reaction pathway with an energy barrier for any conventional model on the B3LYP level of theory, regardless of basis set used, and also independent of surrounding environment (*in vacuo* or PCM). This is consistent with the previous findings of Mauser and coworkers [80]. However, on the HF level, we were able to find reaction barriers, albeit of low height. Table 3.1 summarizes the results for the reaction barriers, reaction enthalpies, zero point energies and isotopic effects of the two conventional model systems "1" and "1b", with different basis sets and with or without a polarized continuum model.

Method	HF/6-31G(d,p)		
Model (atoms)	1 (102)	1b (19)	1 PCM (102)
$\Delta G^{\ddagger}_{12}$	18.415	37.169	20.292
ΔG_{12}	-128.056	-40.346	-112.030
k_{12}	$3.69{\cdot}10^{9}$ s^{-1}	$1.91{\cdot}10^{6}$ s^{-1}	$1.73{\cdot}10^{9}$ s^{-1}
$\Delta G^{\ddagger}_{13}$	18.518	37.290	20.398
ΔG_{13}	-128.088	-40.359	-112.062
k_{13}	$3.54{\cdot}10^{9}$ s^{-1}	$1.82{\cdot}10^{6}$ s^{-1}	$1.66{\cdot}10^{9}$ s^{-1}
α	1.042	1.050	1.043

Table 3.1: Summary of the carboxylation Gibbs free activation and reaction energies (from scaled frequencies) and for different isotopes (^{12}C and ^{13}C) within the conventional model, see Eqs. (3.2) and (3.3). Without and with PCM ($\epsilon = 4.7113$; chloroform) for the 102 atom model. Rate constants k derived thereof are also shown, obtained using eq. (2.67). All energies are in kJ/mol. The isotope effect α is calculated using eq. (3.1).

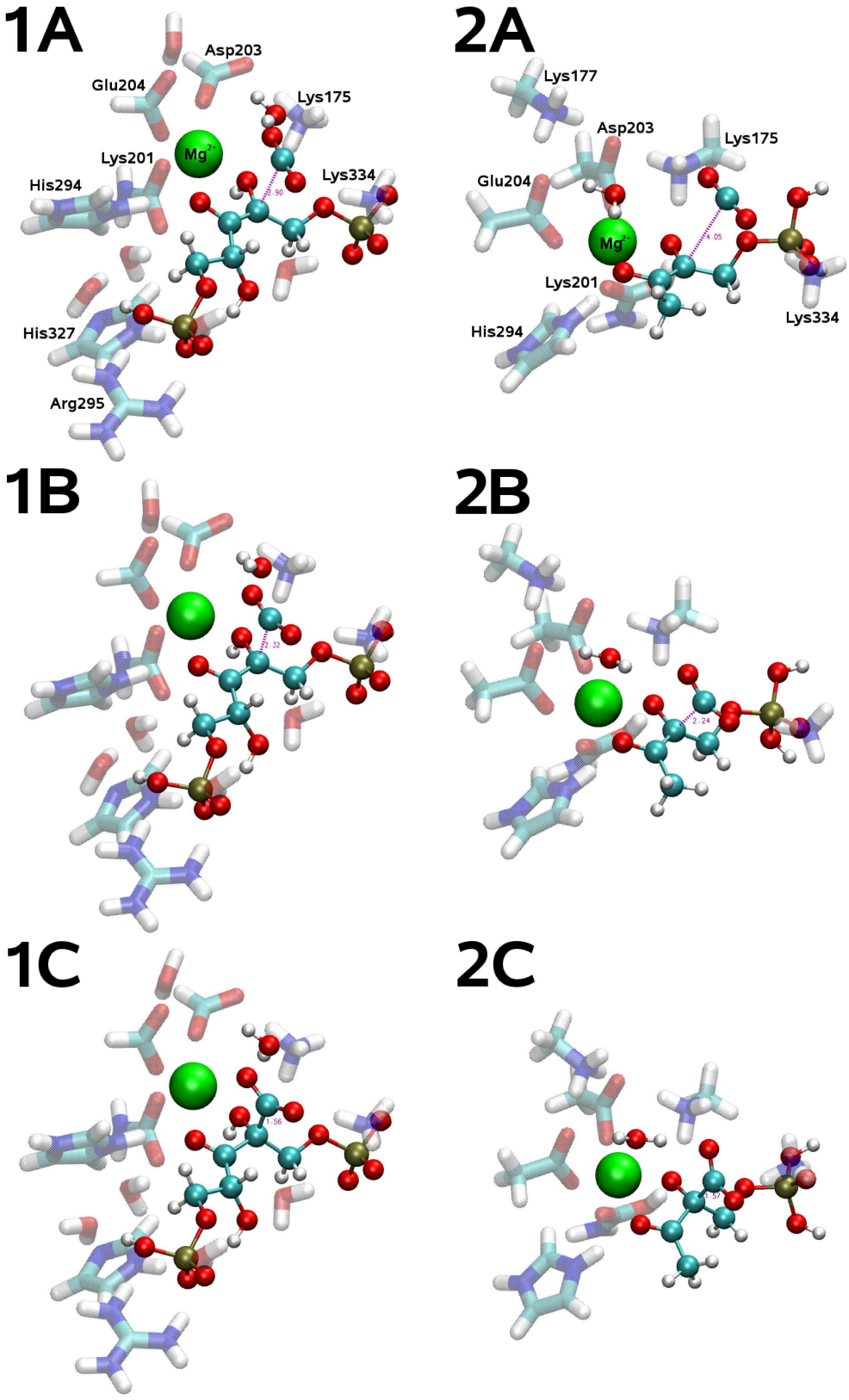

Figure 3.5: "1" and "2" RuBisCO carboxylation models [65, 76], HF/6-31G(d,p); reactant, transition and product state (from top).

3.2.2 New Model

Model Construction

The "new model" is the largest model from the publication of Kannappan and Gready [76]. The B3LYP/6-31G(d,p) *in vacuo* structures were reconstructed from the coordinates delivered in the supplemental files. All other methods and the PCM approaches used the B3LYP/6-31G(d,p) transition state structure as starting point for the scan of the reaction pathway.

The new model consists of CO_2, Mg^{2+}, Lys175, Lys177, Lys201 (carboxylated, "KCX201"), Asp 203, Glu204, His294, Lys334, and a single water molecule bridging the CO_2 to the Mg^{2+} coordination sphere. The substrate molecule, RuBP, was cut between C4 and C5, omitting the P2 group and replacing the C5 by a hydrogen. In contrast to the conventional model, the amino acids of the new model retained a single carbon of their respective aliphatic chain. Otherwise, the build was analogous, and consisted of 77 atoms in total. The resulting model can be seen in fig. 3.5 and will be referred to as model "2" for easier comparison with models 1 and 1b.

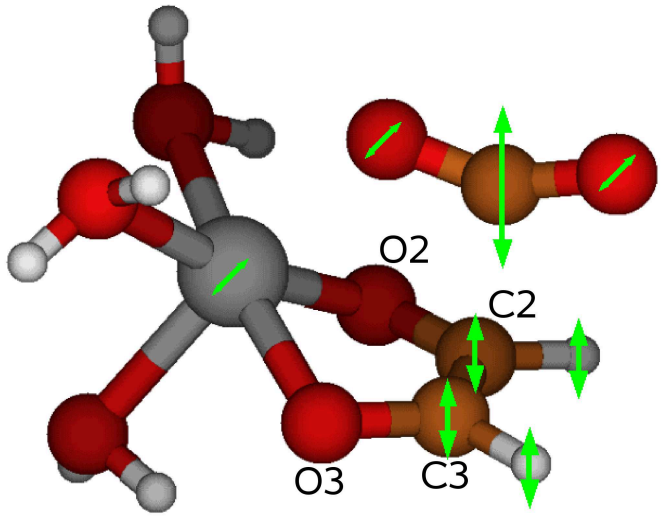

Figure 3.6: Small, conventional RuBisCO carboxylation model 1b, transition state (HF/6-31G(d,p)). Arrows indicate approximate direction and strength of the transition mode.

System treatment

In contrast to models 1 and 1b, no restraints needed to be applied to model 2, as it did not show any unphysical distortions. Any stationary points found were thus a minimum

for all degrees of freedom or a first order transition state.

Results

Our calculations for the ^{12}C isotope reproduced the reaction barrier previously described elsewhere [76]. Table 3.2 contains the results for the model 2, analogous to table 3.1 for the conventional model. Note that the PCM/B3LYP/6-31G(d,p) pathway is the only pathway yielding an endergonic reaction. However, since the backwards reaction does not play a role for the kinetic isotope effect, this is of no further concern.

Method	HF/6-31G(d,p)		B3LYP/6-31G(d,p)	
Environment	Vacuum	PCM	Vacuum	PCM
$\Delta G_{12}^{\ddagger}$	81.062	75.228	39.976	46.873
ΔG_{12}	-34.633	-33.310	-6.335	1.110
k_{12}	$3.91{\cdot}10^{-2}$ s^{-1}	$4.11{\cdot}10^{-1}$ s^{-1}	$6.16{\cdot}10^{5}$ s^{-1}	$3.81{\cdot}10^{4}$ s^{-1}
$\Delta G_{13}^{\ddagger}$	81.180	75.357	40.062	46.991
ΔG_{13}	-34.654	-33.328	-7.063	0.179
k_{13}	$3.72{\cdot}10^{-2}$ s^{-1}	$3.90{\cdot}10^{-1}$ s^{-1}	$5.95{\cdot}10^{5}$ s^{-1}	$3.64{\cdot}10^{4}$ s^{-1}
α	1.049	1.053	1.036	1.049

Table 3.2: Summary of the carboxylation Gibbs free activation and reaction energies (from scaled frequencies) and for different isotopes (^{12}C and ^{13}C) within model 2, see Eqs. (3.2) and (3.3). Without and with PCM ($\epsilon = 4.7113$; chloroform). Rate constants k derived thereof are also shown, obtained using eq. (2.67). All energies are in kJ/mol. The isotope effect α is calculated using eq. (3.1).

We also performed calculations with the smaller basis set 6-31G(d) for model 2, recalculating the vacuum models with both HF and B3LYP (see appendix A.1). It was found that the reaction barriers were lowered by about 2.5 kJ/mol, but reaction enthalpies were unaffected and the isotopic effect changed only by 0.003 or less.

3.3 Discussion and Summary (RuBisCO)

Regardless of model and method used, the isotope effect calculated is in the range of the experimental value, which is well reproduced. Our calculated values range from 1.036 to 1.053, while the experiment shows an effect of about 1.03 [18].

Thus, the occurrence of the isotope effect is not controlled by the mechanism or environment. The exact magnitude of the effect however can be expected to vary depending on the employed protein environment. This is consistent with the observation that RuBis-

COs from different organisms vary slightly in the extent of their respective isotope effect [64].

We calculated the isotope effect only for the carboxylation reaction step. It can be seen that it always slightly exceeds the experimental, overall reaction isotope effect [18]. This is reasonable, as the total rate constant $k_{tot,i}$ for an isotope i can be approximated as

$$k_{tot,i} = \left(\frac{1}{k_{bind.,i}} + \frac{1}{k_{enol.,i}} + \frac{1}{k_{carb.,i}} + \frac{1}{k_{hydr.,i}} + \frac{1}{k_{c-c\ spl.,i}} + \frac{1}{k_{rel.,i}} \right)^{-1} \tag{3.4}$$

with different rates k for each indexed process. The binding (*bind.*) and release (*rel.*) of reactants and products are essentially diffusion controlled, and thus may have an influence on isotope discrimination, especially the binding of CO_2. The other rate constants are analogous to the reaction pathway given in fig. 3.1.
As long as all partial reaction steps in (3.4) are less critical for the isotope effect than $k_{carb.}$, we see that with (3.1),

$$\frac{k_{tot,12}}{k_{tot,13}} < \frac{k_{carb.,12}}{k_{carb.,13}} \quad . \tag{3.5}$$

Thus the α calculated purely for carboxylation step has to yield a higher value than the total reaction. In both models this effect can be seen, although to a slightly different extent.

Assuming the carboxylation step to be rate limiting, as suspected in some sources [83], the reaction barriers of the new model are more reasonable than the ones of the conventional model. Taking the experiments on spinach RuBisCO into account, one would expect a barrier of 71.6 kJ/mol (from (2.67) for a $k_{carb.} \approx k_{tot}$ of 1.75 s^{-1} [72]). For the conventional models treated by Hartree-Fock, the barriers are much lower than expected, regardless of employed model treatment.
For the new model, calculated Hartree-Fock barriers are higher than experimental values. This could be explained with missing correlation energy at the transition state, since it is well known that HF activation energies are typically too large. B3LYP barriers are in contrast slightly too low compared to experiment which is also in agreement with general experience. In summary, the new model seems to be a better one under the given assumptions. However, other sources indicate that carboxylation is spontaneous and the rate limiting step is further down the reaction pathway [65, 64]. It might thus as well be

possible that several reaction barriers of similar size contribute to the total rate constant, or maybe even cooperative effects between the two active sites of a 2L RuBisCO complex occur.

PCM treatment of all models shifts the reaction barriers to values closer to experiment, but no systematic behavior can be observed. Here, the new model is probably more susceptible to errors: Most of the conventional model is frozen, while the geometry of the new model is allowed to change more drastically. It is not surprising that the effect of the PCM on the new model is thus larger in comparison to the vacuum models. This could be the reason why the PCM/B3LYP pathway of the new model is endergonic in contrast to the other calculations. With the given data, PCM cannot be directly considered to be a general improvement to the model.

The origin of the isotopic effect can be located easily by taking a look at the raw data (see appendix A.1). Most of the energetic differences between the ^{12}C and ^{13}C reaction barriers arise from the change in ZPE alone. Already at 0 K, by simply adding the ZPE correction to the electronic energy, most of the energy barrier difference occurs. For example, the new model with B3LYP/6-31G(d,p) yields a barrier difference of 87 J/mol between ^{12}C and ^{13}C. For the ^{12}C isotope, the ZPE rises from the reactant to the transition state by 5461 J/mol, while for the ^{13}C isotope, it only rises by 5535 J/mol, yielding a difference of already 74 J/mol in favor of ^{12}C. This way, it can be seen that effects arising from higher temperature do not play a major role.

As the carboxylation process does not create large geometric distortions in either model 1 or 2 (even though enforced by restraints in the conventional model), and considering that the ZPE contributes for the most part of the isotopic effect, it is not surprising that both models give similar results. Both models use the same reaction coordinate, simply following of the C-C-distance as reaction pathway (see fig. 3.6). From there it is seen that the mode corresponding to the imaginary frequency at the transition state has a strong contribution of the CO_2 bending mode which is required to form the carboxylated product. It can be suspected that regardless of the environmental details, the isotopic

	${\omega_b}^{12}$	${\omega_b}^{13}$	$\frac{{\omega_b}^{12}-{\omega_b}^{13}}{2}$	$e^{\frac{\hbar(\omega_b^{12}-\omega_b^{13})}{2k_BT}}$	$1+\frac{\hbar(\omega_b^{12}-\omega_b^{13})}{2k_BT}$	$\frac{\omega_b^{12}}{\omega_b^{13}}$
HF/6-31G(d,p)	684.48	665.00	9.74	1.047	1.046	1.029
B3LYP/6-31G(d,p)	629.40	611.49	8.95	1.044	1.043	1.029

Table 3.3: Scaled vibrational frequencies of the CO_2 bending mode, given in cm^{-1}. Quantities derived thereof, in particular the isotope effects calculated from various approximations, are also shown.

effect should be well reproduced by any model featuring this specific reaction coordinate. To further support this, one can simply compare the bending modes of CO_2 *in vacuo*. One finds that already a simplistic assessment based of the CO_2 bending mode compares well to the isotopic effect (see table 3.3). Thus, the simple bending "speed" of the CO_2 delivers a simple, albeit not fully accurate, picture of the isotope effect's origin.

A more accurate approximation is to include the ZPE difference of the CO_2 bending mode into (2.67), as the reaction coordinate is strongly dependent on this mode. If we define the isotopic effect only on the basis of ZPE difference of the bending mode, we get

$$\alpha = e^{(\Delta G_{13}^{\ddagger}-\Delta G_{12}^{\ddagger})/k_BT} \approx e^{\hbar(\omega_b^{12}-\omega_b^{13})/2k_BT} \tag{3.6}$$

with $\hbar(\omega_b^{12}-\omega_b^{13})/2$ being the ZPE difference of the $^{13}CO_2$ and $^{12}CO_2$ isotopomers in the reacting (bending) mode. Since $k_BT >> \hbar(\omega_b^{12}-\omega_b^{13})$ at room temperature, we can further approximate

$$\alpha \approx 1+\frac{\hbar(\omega_b^{12}-\omega_b^{13})}{2k_BT} \quad . \tag{3.7}$$

(3.6) and (3.7) omit all influences of the environment or of the RuBP, and thus enables us simply to take CO_2 *in vacuo* as smallest possible model system. The approximated isotopic effects can be found in tab. 3.3 and are always in the range of the other isotope effects calculated above for the active site models.

In conclusion, we suggest that the isotopic effect arises from differences in the frequency of the CO_2 bending vibration for the two isotopomers. The bending mode of CO_2 dominates the atom motion at the transition state along the way towards the carboxylated ribulose. In Eyring transition theory picture, this leads to a loss of zero-point energy of the reactant configuration for $^{13}CO_2$, and thus to an effectively slightly larger activation free energy and smaller carboxylation rate. In the simpler Arrhenius picture, the isotope effect can be traced back to the smaller attempt frequency of the reactive

bending vibration of $^{13}CO_2$. These assumptions hold if the carboxylation step is the most important one regarding the isotope effect.

The findings of this chapter suggest that the isotopic effect can be acceptably predicted and explained by any kind of model containing a realistic C-C-distance reaction coordinate. Altogether, we can assume that the isotope effect is based on the properties of the CO_2 reaction coordinate alone, and is only slightly modified by environmental details. Both the "conventional" and the "new" model produce an acceptable isotope effect. Thus, from the isotope effect alone, we cannot decide which of the two models is more realistic.

Thermal protein dynamics may increase the accuracy of the results. The binding and release dynamics may also be included in this manner. Future studies, incorporating QM/MM approaches of larger RuBisCO portions, may elucidate the reaction process even further, aiming at finding systematic relationships between protein structures and the isotopic effect.

Chapter 4

BLUF Domain Characterization

4.1 Blue-light Receptor Domains

Plant photoreceptors play a crucial role for any photosynthetic organism in the adjustment of growth direction. More immediate responses, such as the shutdown of the photosynthetic apparatus under light stress, have also been reported [84, 85]. Several classes of photoreceptors exist, and the basis for all of them are chromophore molecules embedded in a protein environment [86].
The class of *b*lue *l*ight sensors *u*sing *f*lavin (BLUF) protein domains is a family of photoreceptors with flavin-based chromophores [19]. It is not known how photo-excitation of BLUF translates into a signal recognizable by the living cell. The signal is suspected to be of conformational origin, but the exact pathway from an excited flavin chromophore to a conformationally changed BLUF domain remains unexplained.

In this chapter, the spectral characterization of BLUF domains *via* time-dependent density functional theory (TD-DFT) will be presented. Our aim is to elucidate the connections between the chromophore and the immediate protein environment. After a short, general introduction on BLUF domains (this section), we will continue with the presentation of the models built for BLUF domains (sec. 4.2). This will also include dynamic propagations of BLUF domains in aqueous solution using classical force fields (sec. 4.3). Finally, the results of the calculations based on these models and trajectories will be discussed in section 4.4.

4.1.1 BLUF domain properties

BLUF domains are small photoreceptive protein domains consisting of about 120 amino acids (about 20 kDa) with a flavin chromophore. Their absorption spectrum shows the typical flavin spectrum slightly shifted to higher wavelengths. This shift is independent of the flavin species in the protein (riboflavin, Rf; flavin mononucleotide, FMN; flavin-adenine dinucleotide, FAD; see fig. 4.1) [87, 88]. The secondary structure is of a common $\beta\alpha\beta\beta\alpha\beta\beta$ motif. It is extended sometimes by two additional α helix regions, dependent on the type of BLUF domain. The overall folding of the protein forms a sheath in which the flavin chromophore is embedded, isolated from the solvent.

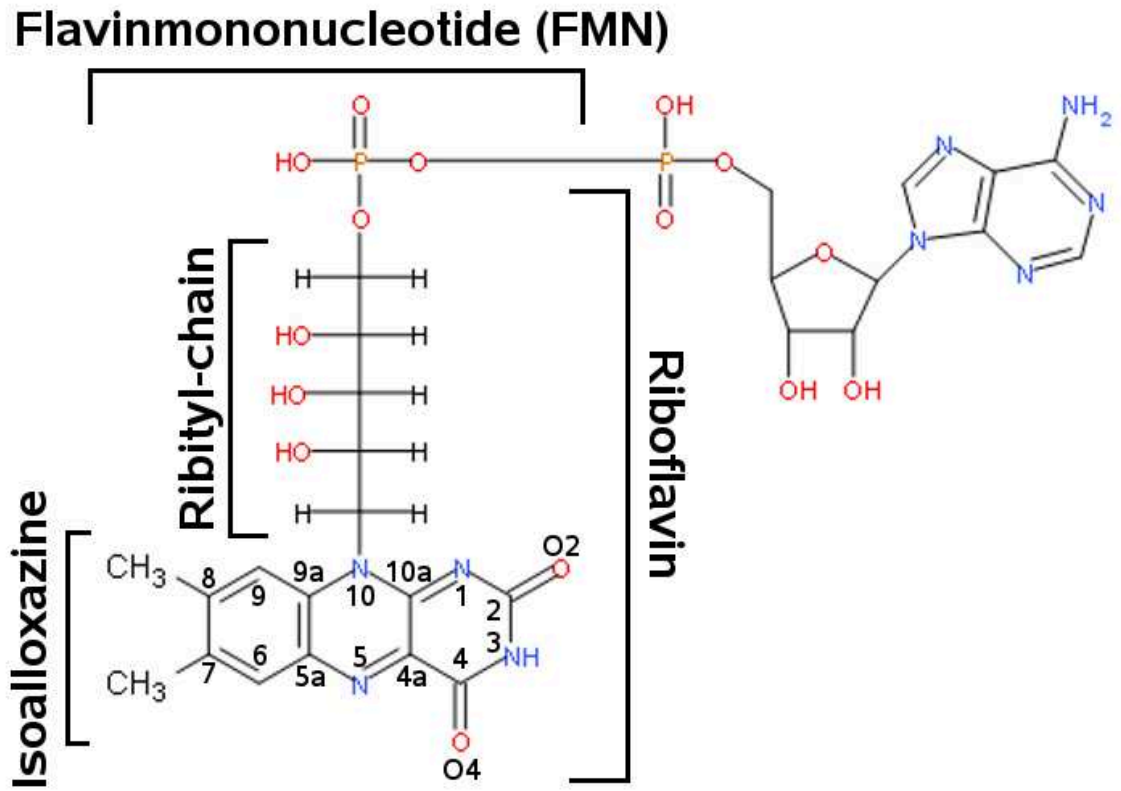

Figure 4.1: Flavindinucleotide (FAD) and derived types of smaller flavin chromophores (and ribityl chain for clarification). Cut bonds are saturated with hydrogens to create the respective smaller species. Numbering only for the isoalloxazine ring due to its importance for the BLUF photocycle.

While BLUF domains share general construction principles, it is not known if all BLUF domains share the same functional mechanism [19]. From a conceptional perspective, they share a lot of their basic construction with the LOV type domains (light-oxygen-voltage

[89, 90, 91]). Both types of photoreceptors are very small: They basically only shield the chromophore from the solvent. Thus, it is generally assumed that their sole purpose is to transfer the light signal received to a conformational change affecting a C-terminal effector domain [19].

BLUF domains have been identified in a variety of proteins. BLUF was found responsible for the regulation of, *e.g.*, cyclases [19] or transcription factors [84]. Interestingly, BLUF was also identified to exist on its own, having no effector domain at all [92].
The functional aspect of BLUF is generally assumed to be exclusively photo-regulatory. However, due to the obvious similarity to the LOV domains, one should not be surprised if other stimuli may be found to work with BLUF, too [84, 85]. In this work we will concentrate on the photo-regulatory aspect regardless.

The main challenge for theory with respect to all possible photoreceptor domains is the identification of both the "resting" and "signaling states and their connecting pathways. These states are often called *dark* and *light* states, but also *blue* and *red* for reasons explained below. To avoid confusion with spectroscopic features in the upcoming discussion, we will stick to the "resting/signaling notation. For flavins in general, and for both the LOV and BLUF domains, the assignment of the two states was done differently in several studies, none of them fully answering all questions so far [89, 90, 91, 93, 94, 95, 96, 97, 98]. A proposed photocycle on the basis of experimental findings for the BLUF domains is depicted in figure 4.2 for the case of AppA[1].

In contrast to the LOV domains, the NMR and some crystal structures for the BLUF domains [100, 101] revealed differences in positioning of certain amino acids [102, 103, 104, 105]. These are suspected to be related to the BLUF functional mechanism. It became a heated debate as to which conformation is the resting state of BLUF domains. Yet, it also hinted towards which amino acids might be involved in transduction of the light signal to the conformational change. The proposed fact that BLUF domains do

[1]AppA is an abbreviation for "activation of photopigment and puc expression antirepressor". It is a purely descriptive term, denoting that AppA binds to a repressor corresponding to, *e.g.*, expression of light harvesting complexes [84].

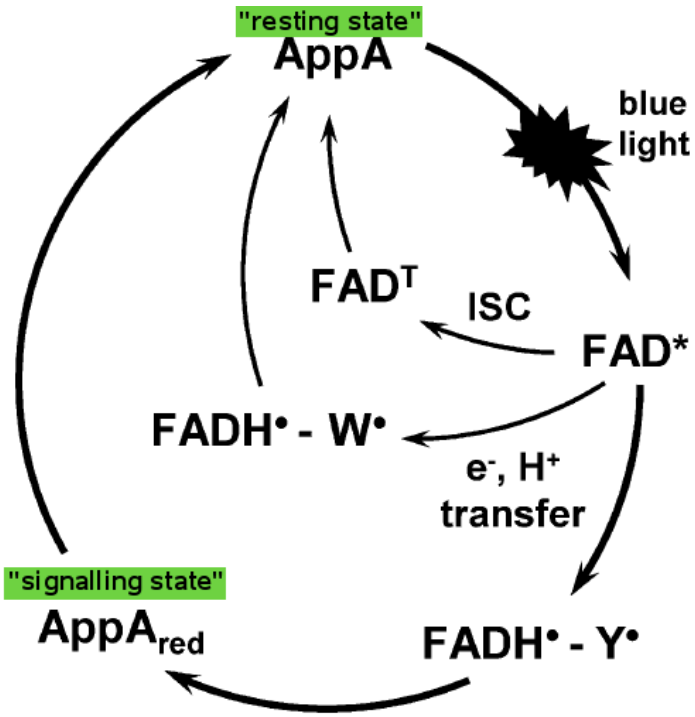

Figure 4.2: AppA BLUF photocycle, adapted from [99]. FAD*: singlet excited state, FAD^T: triplet excited state, ISC: inter-system crossing, FADH$^\cdot$: radical intermediate after hydrogen transfer from the associated amino acid. W and Y are one-letter notations for tryptophane and tyrosine, respectively. For the definition of the term AppA, see text.

not change their chemical constitution upon excitation was however problematic: It was difficult to characterize the reaction of the flavin binding site to an electronic excitation due to the lack of a "real" chemical difference between the resting and the signaling state [92, 106, 107].

In the next section, we display which structural data are available for the building of a viable model. We will however see that this data is not suited to define a single, unambiguous geometry of the two states in BLUF domains.

4.1.2 Structure/function assignment

As stated above, the structural data basis for the BLUF domain is large, but ambivalent. The corresponding analysis is hampered by the fact that not all BLUF domains have the residues conserved which are suspected to be of functional importance. Most notably in this respect is Trp104, which is not present in some BLUF domains such as, *e.g.*, BlrP1

[108]. As such, we will from now on implicitly refer to BLUF domains containing the described amino acids. The other types of BLUF domains may or may not differ in mechanism, although we would like to point out that we are discussing the most common type of BLUF domain here. Numbering of residues will be as in the AppA BLUF domain analyzed by Grinstead and coworkers [100].

Since BLUF domains are small compared to other protein structures like enzymes, NMR studies of these photoreceptors are possible [100, 109]. We have therefore two sources of structural data available which allow for cross checking between the NMR and the crystal structures, thus sorting out possible artifacts. A comparison of the available structural data based on the two methods displays differences in certain structural details. The predominant difference, the positioning of Trp104 and Met106 (shown in fig. 4.11 further below in sec. 4.3), resulted in the dispute within the community, which we already mentioned above. Some researchers believe that the exchange of positions of these two residues is the primary geometrical change after photo-excitation. They think that this is the basis for signal transduction to the surface, although this again is challenged by others [110]. It is however unclear which of these Trp/Met positions corresponds to the resting and which to the signaling state.
Since we based our study on an ensemble of different structures, the structural assignment is mostly irrelevant for our purposes. Yet, it will become clear that based on our results [96, 111], we can contribute strong arguments favoring one of the two sides.

Apart from the ambiguous structural data, there are several other sources to distinguish the resting from the signaling state. First, a 10 nm redshift in the UV/vis spectrum of the BLUF domain occurs upon formation of the signaling state for the energetically lowest maximum ($S_{0\to1}$ transition). There is also a not as strong shift for the second maximum ($S_{0\to2}$) [112]. An example for this shift is given in fig. 4.3. The resting state exhibits a $S_{0\to1}^{\text{resting}}$ maximum at about 445 nm, which is red-shifted by 10-15 nm in the signaling state (approx. 0.10 eV), with a loss in intensity. The $S_{0\to2}^{\text{resting}}$ maximum, at about 365 nm, shifts by 5-10 nm to the red (approx. 0.05 eV), slightly gaining intensity. Both shifts are thus in the same order of magnitude, but rather small, and exact values may

vary depending on the type of BLUF domain. The details on both the $S_{0\to1}$ and $S_{0\to2}$ bands are given in table 4.1.

Band	Exp. λ_{max}/nm (eV)	Transition
$S_{0\to1}^{\text{resting}}$	445 (2.79)	$\pi \to \pi^*$
$S_{0\to2}^{\text{resting}}$	365 (3.40)	$n \to \pi^*$
$S_{0\to1}^{\text{signaling}}$	460 (2.70)	$\pi \to \pi^*$
$S_{0\to2}^{\text{signaling}}$	371 (3.34)	$n \to \pi^*$

Table 4.1: Details on the low-lying flavin (in AppA) UV/vis maxima, assembled from [122, 91]. Note that we only refer to optically allowed bands for the band terminology.

Other significant shifts are seen in the IR spectrum: Without going too much into detail, FTIR studies indicate that the hydrogen bonding to the O2 and O4 atoms of the chromophore is strengthened upon signaling state formation [112, 92]. In turn, the force constant between C2=O2 and C4=O4 is lowered (for atom numbering, see fig. 4.1). Especially the latter is of interest, as O4 is next to potentially important residues, such as Asn44, Gln63 and Trp104 (see below).

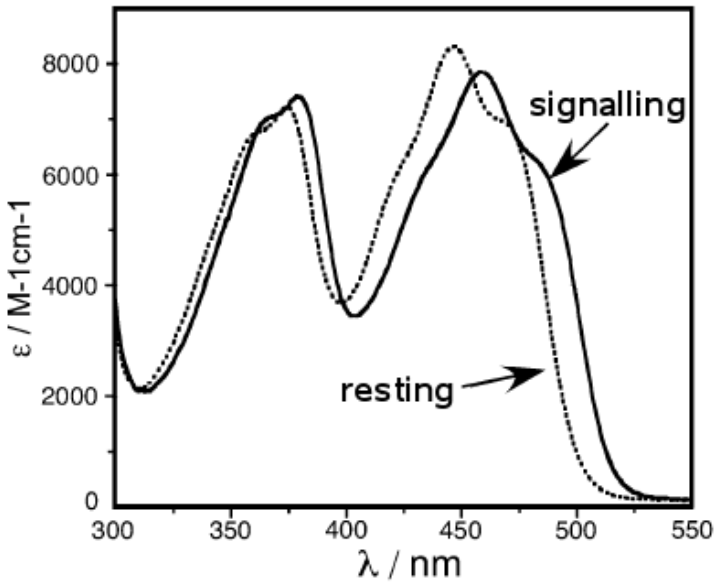

Figure 4.3: Experimental UV/vis absorption spectra of resting and signaling state for the AppA BLUF domain, adapted from [122]. Note that the spectra of AppA and BlrB (see sec. 4.3) BLUF domains are virtually identical.

Since we will be working with non-QM-optimized structures from NMR and MD, we cannot derive any information regarding the IR spectrum; the necessary harmonic approx-

imation would require conformational minima[2]. We will however explore the influence of the flavin binding pocket geometry on the UV/vis transitions.

Before we can construct the models, we have to discuss the possible functionalities of amino acids close to the chromophore. This will be done in the following section.

4.1.3 Roles of single amino acids

It is necessary to discuss the immediate surroundings of the flavin chromophore in more detail before building a model. This can be done regardless of which structure is the resting and which is the signaling state. While Trp104 or Met106 may or may not be close to the chromophore, other amino acids are always close to the active region. Most prominently, Tyr21 and Gln63 have been found to critically affect the photocycle: No photocycle occurs once these residues are replaced by alanine [113]. These two amino acids remain close to the chromophore, independent of Trp104/Met106 positioning.
It is suspected that initial photo-excitation leads to a charge transfer from Tyr21 to the flavin. This in turn should result in a conformational change for Gln63. Other sources discuss a possible tautomerism of the Gln63 hydrogens, leading to a high-energy construct of an enolized Gln63 [114, 93]. Yet, our own test calculations (not shown) yielded energy differences between the standard Gln63 and its enolized form of about 70 kJ/mol, strongly disfavoring this mechanism. We will therefore refer to the more stable non-enolized Gln63 only, as for our study, the exact Gln63 events are of no concern.

Other amino acids close to the chromophore include Asn45, which is reported to be important for the binding of flavin, but not for the photocycle. His44, which may be Asn or Arg in some BLUF domains, seems to provide a positive charge for the polar part of the flavin's isoalloxazine ring, specifically to the O2 atom. It seems to be of no importance for the photocycle, too [103].

[2]Test calculations have shown that the resulting IR spectra display mostly real frequencies and the resulting bands are in the regions of the experimental spectra. However, without further sampling and comparison to optimized models, the reliability of the NMR-based IR spectra cannot be evaluated.

A detailed picture of the general arrangement of the flavin binding pocket can be seen in fig. 4.4. Depicted amino acids, which were not mentioned up to now, will be discussed later on. Their possible role will become evident throughout the course of our investigations. As we are interested in building a model with which we can qualitatively reproduce the $S_{0\to1}$ transition for both the resting and signaling states, we will now discuss the construction of an appropriate model.

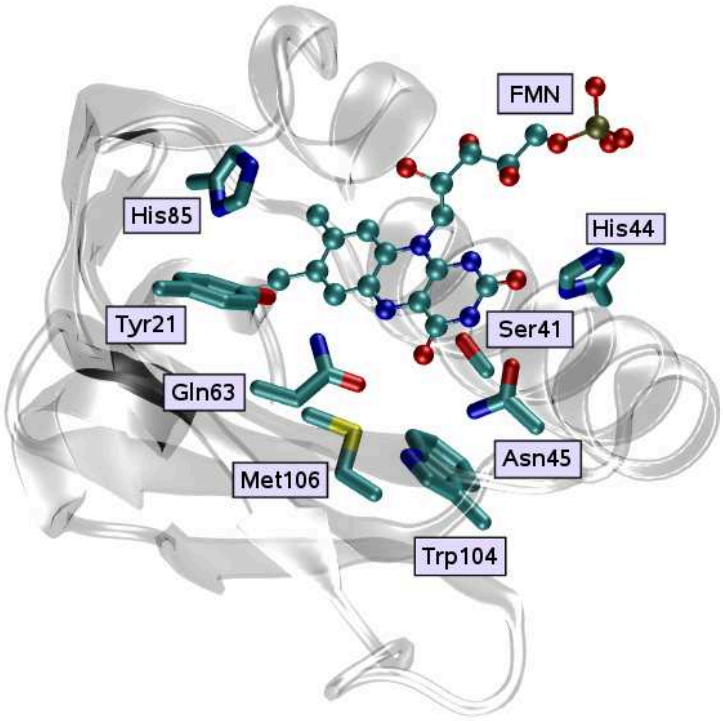

Figure 4.4: Geometry of the AppA BLUF flavin binding pocket with selected non-aliphatic amino acids in close contact to the isoalloxazine ring. Structural data taken from [101]. α-helix covering the FMN partially deleted for better display.

4.2 Spectral Characterization

To make sure that we can grasp a wide variety of possible BLUF structures, we based our model on a NMR structure (PDB code 2BUN, AppA BLUF [100]). NMR structures have the advantage of delivering 20 varying structures instead of a few (mostly identical) proteins in the crystal's unit cell. Furthermore, as we want to make sure that the BLUF domain is in a native conformation, small proteins are more likely to be conformationally influenced by their crystal neighbors because their volume to surface ratio is much smaller. Finally, the NMR structure was taken at room temperature and in solution, resembling

its natural conditions. NMR, however, can also contain errors based on the high mobility of the atoms. For the reproduction of the NMR signals by the structural model, several NMR structures have to be constructed and put together to form an overall signal. In this process, artifacts of various kinds may occur, as structural features may be mixed unnaturally.

For all calculations of this section, the *Gaussian03* package [40] was used.

4.2.1 Model construction

Since we cannot know the amino acids important for the $S_{0\to1}$ transition beforehand, we have to account for all possible candidates close to the chromophore. As the HOMO/LUMO transition is mostly localized at the isoalloxazine ring, we can omit the contacts to the ribityl chain. Aliphatic amino acids, too, will likely have only minor influence, and are also neglected. This still leaves us with a set of amino acids which is different for each of the 20 NMR structures. To ensure comparability, we have to include the amino acids in all calculations, even if they are only close in one of the 20 structures.

This model consisted of a Rf as chromophore (constructed from a FMN in the NMR structure), Tyr21, Ser41, His44, Asn45, Gln63, Met76, His85, Trp104 and Met106. The amino acids were cut in the manner that the connection to the backbone was represented by a single methyl group. Met76 and His85 were later found to be spectrally inactive, but remained included for purposes of comparability. After optimizing the hydrogens with B3LYP/6-31G(d), TD-B3LYP/6-31G(d) calculations for 30 singlet excited states were performed to get an overview of the contributions to the HOMO as well as the LUMO Kohn-Sham orbitals for the $S_{0\to1}$ transition (drawn at a contour value of 0.02 using the *molden* program [115], see fig. 4.7). It was found that the LUMO is exclusively located at the isoalloxazine ring, regardless of the NMR structure. In contrast to the LUMO, the HOMO was found to be partially extending to the amino acids: Strong contributions were found on Ser41, Asn45, Gln63; rarely Tyr21 contributed, too.
As a result, we extended our model slightly to gain a more complete environment for

the simulation of the HOMO; and we included two more structures: Close to the Tyr21, there is a serine residue (Ser23), which could possibly modify the properties of the Tyr21 HOMO contribution. We also included the C_α of Ile37 and the peptide bond (OCNH) to Val38, filling missing valences with hydrogens. This backbone structure's oxygen was found to be a possible hydrogen bonding partner in about 1/3 of the NMR structures for Ser41. For all the other HOMO-contributing amino acids, no non-aliphatic structures were close by, for which we may have suspected an influence on the HOMO and thus on the $S_{0\to1}$ transition. The full model consisted of 47 Rf atoms and 118 protein atoms, 165 in total. Fig. 4.5 is a representation of the full model in one of the NMR geometries.

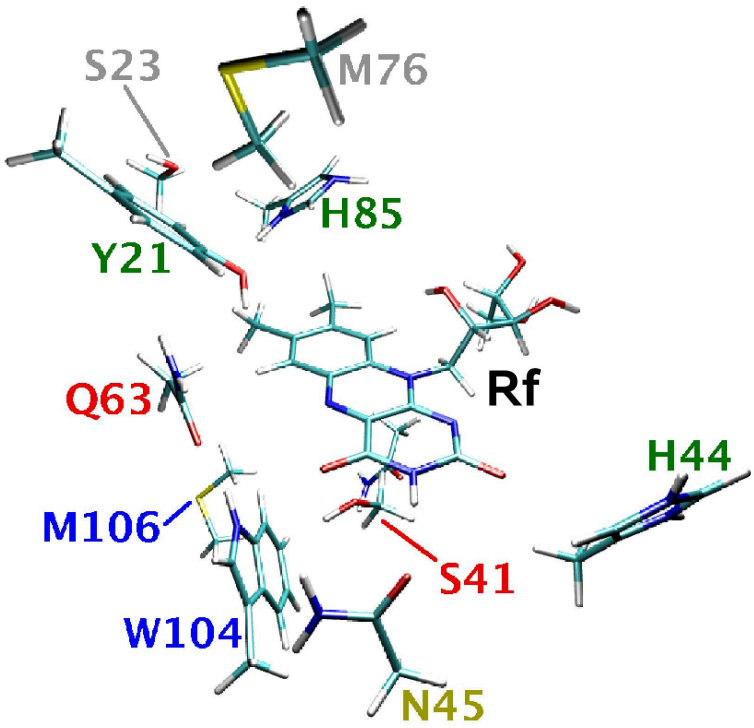

Figure 4.5: Full BLUF model composition, single letter notation used for labels, numbering as in PDB dataset 2BUN [100] (AppA BLUF). Ile37/Val38 backbone segment unlabeled, violet, below riboflavin (Rf). Colors of labels indicate structural flexibility in the NMR structure data and HOMO contribution. Grey: No HOMO contribution, inflexible. Blue: No HOMO contribution, small flexibility. Green: Weak HOMO contribution, inflexible. Yellow: Strong HOMO contribution, inflexible. Red: Strong HOMO contributions, highly flexible.

4.2.2 Spectral analysis

The full model was optimized in terms of hydrogen atoms only, on the B3LYP/6-31G(d) level. A full optimization was neither possible due to the system's cut-off border, nor desired, as we did no want to loose the implicit thermal ensemble of the NMR struc-

ture. The 20 partially optimized NMR structure models were then subjected to TD-B3LYP/6-31G(d) calculations, requesting 50 singlet states. The resulting line spectra were broadened by Gaussians to account for thermal movement, finite lifetimes and incoherent broadening:

$$\epsilon(\lambda) = C \sum_I f_I e^{-(1/2)\left(\frac{(1/\lambda)-(1/\lambda_I)}{\sigma}\right)^2} \quad . \tag{4.1}$$

Here, f_I is the oscillator strength of the I-th transition and λ_I its wavelength. C= $1/(\sigma\sqrt{2\pi})$ is a normalization factor, and σ is the Gaussian width which we chose as 750 cm^{-1} throughout the simulations. Intensities were then converted into extinction coefficients ϵ by multiplying with $4m_e c^2 \epsilon_0 ln(10)/(Le^2)$ ($= 4.318998 \times 10^{-10}$mol/m), where m_e is the electron mass, c the speed of light, ϵ_0 the dielectric constant of the vacuum, L Avogadro's constant and e the elementary charge. The TD-B3LYP/6-31G(d) method was shown to perform well for the low-lying $\pi \rightarrow \pi^*$ (or $n \rightarrow \pi^*$) transitions which dominate the UV/vis spectrum of LOV and BLUF domains [89, 90, 91].

It is clear that despite the broadening, our results can only be qualitative in nature. This is due to the fact that vertical absorption spectra necessarily show wavelengths higher or equal to the true vibronic $0 \rightarrow 0$ transition. Additionally, the vibrational fine structure observed in the experimental spectrum (fig. 4.3) will not be reproduced by the line spectra gained from a TD-DFT calculation [116, 117].

All average spectra

The spectrum resulting from taking the arithmetic average over all 20 NMR based UV/vis spectra for the full model resembles the $S_{0\rightarrow1}$ transition well. It is however shifted to blue for the $S_{0\rightarrow2}$ transition (see fig. 4.6, solid line). To make sure that the model is also valid for other sets of geometries, we conducted a short 100 ps MD for each of the NMR structures (Charmm force field [118] employing the Tinker package [48], NVT setup with Berendsen thermostat [57], solvated in about 4100 TIP3P water [119] molecules using a cubic box with an edge vector length of 52.0176 Å; FMN parameters taken from [120]). We took snapshots every 20 ps, and thus increased our previous set of 20 structures by 20 x 5 geometries. The snapshots were treated in the same way as the original NMR structures (except for calculating 90 singlet states this time). Of these 120 structures in total, 6 encountered SCF convergence problems during the treatment and were omitted.

The line spectra from the remaining 114 geometries were broadened as described above and averaged. The resulting spectrum can be seen in fig. 4.6 (dashed line).

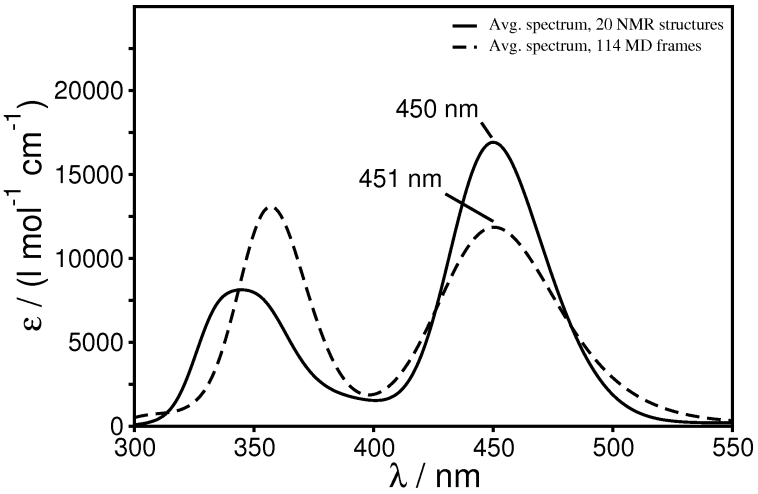

Figure 4.6: Full model average TD-B3LYP/6-31G(d) spectra (50 singlet states), calculated line spectra broadened by (4.1). Solid graph spectrum is based on 20 NMR structures, dashed graph spectrum is based on 114 Charmm trajectory snapshots (see text for details). Maxima of $S_{0\to1}$ transition labeled.

It can be seen that the full model is capable of reproducing the $S_{0\to1}$ transition for the MD based structures, too. The calculated maximum is close to the experimental resting state maximum of 445 nm. It is however not advisable to use this as any kind of proof for the distinction between the two states, resting or signaling. The signaling state maximum of 460 nm is close enough to the calculated value to be in the method's error range. Thus, the NMR structures could represent either the resting or the signaling state (or *both*, given the deficiencies of NMR), and we cannot distinguish between them using the present data. The amino acids included in the model seem however to be already enough for the reproduction of the $S_{0\to1}$ transition. While the intensity of the $S_{0\to1}$ transition still changed for the MD structures compared to the original 20 NMR geometries, the position of the maximum was unaffected.

The calculated $S_{0\to2}$ transition is also well reproduced. This is surprising, as to the TD-B3LYP error which is reported to increase for higher transitions [121]. The large change of the $S_{0\to2}$ maximum in the 114 MD snapshot average spectrum may however be a hint for a greater dependence of the sampling of the environment for $S_{0\to2}$ in contrast to $S_{0\to1}$.[3]

[3]We will see in chapter 5 that this also holds true for other systems.

We calculated the average maxima and oscillator strengths throughout the NMR geometry set. The results are summarized in table 4.2; however, note that the calculated $S_{0\to1}$ and $S_{0\to2}$ bands may be composed of several transitions, depending on the analyzed geometry. We used only the main transitions with the highest corresponding oscillator strengths in the respective spectral regions for the analysis in table 4.2. The mean values in table 4.2 thus differ slightly from the ones seen in fig. 4.6. This was not performed for the MD structures, as the variance in the shape and order of the orbitals did not allow for any systematic analysis, not even with the oscillator strength criterion used for the NMR geometries.

Band (Set)	Avg. λ_{max} /nm (eV)	Mean Dev.(λ_{max}) /nm (eV)	Avg. f_I	Mean Dev.(f_I)
$S_{0\to1}^{NMR}$	452 (2.74)	12 (0.07)	0.1373	0.0257
$S_{0\to2}^{NMR}$	360 (3.45)	18 (0.18)	0.0621	0.0309

Table 4.2: Average values of the TD-B3LYP/6-31G(d) calculated low-lying flavin (in AppA) UV/vis transitions based on the NMR geometries. Labeling as in table 4.1.

We have established that our model is reliable for the $S_{0\to1}$ transition, maybe also for $S_{0\to2}$, at least when taking the NMR geometry set. We will now see if there are any connections between structural and spectral features.

4.2.3 Structural analysis

An overview on the dynamic structural features of the full model is given in fig. 4.5. Two amino acids have been found to strongly contribute to the HOMO of the S_1 state (see fig. 4.7) and are highly flexible regarding their position throughout the NMR dataset: For Gln63 there have already been reports on its importance in the BLUF photocycle [103, 113]. The other, Ser41, has not been experimentally addressed before in terms of spectral importance and functionality. We therefore aimed at predicting the possible influences of Ser41 positions on the BLUF spectrum, together with exploring the same features for Gln63 within our model system.

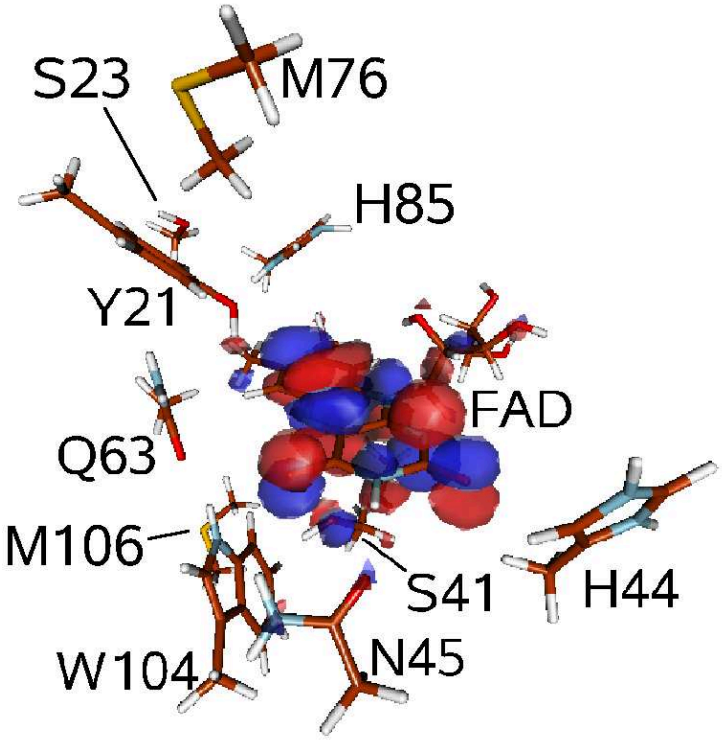

Figure 4.7: Full model B3LYP/6-31G(d) showing the highest occupied Kohn-Sham orbital with significant coefficients at FAD, participating in $S_{0\rightarrow1}$ for 2BUN9 substructure 2, from [96]. The orbital has high coefficients on the Ser41 OH-oxygen.

Looking for a systematic influence on the chromophore, we screened the 20 NMR structures for distinct positions of both Gln63 and Ser41. As geometrical features always change slightly, we cannot define distinct positions for any of the amino acids. However, if we can concentrate on the presence/absence of hydrogen bonds, we can still assign the structures to different groups. In the case of Gln63, it has been proposed that the hydrogen bond between Gln63 and Trp104 may break/form upon forming the signaling state [104, 113]. Thus, we decided to split the 20 NMR structures into two groups, one group having the Gln63-Trp104 hydrogen bond ("Q_{formed}"), the other lacking this feature ("Q_{broken}"). A pictorial representation of the two positions is given in fig. 4.8A.

As there were no reports for Ser41 so far, we had to come up with a grouping of our own. We found that Ser41 in some of the NMR structures was turned away from the isoalloxazine ring, forming a hydrogen bond with the backbone peptide bond oxygen between Ile37 and Val38 (in fig. 4.5 unlabeled, behind Rf; or fig. 4.8B). We termed these groups "S_{flav}" or "S_{back}", depending on whether the Ser41 is facing the flavin or the backbone, respectively. The two situations can be seen depicted in fig. 4.8B.

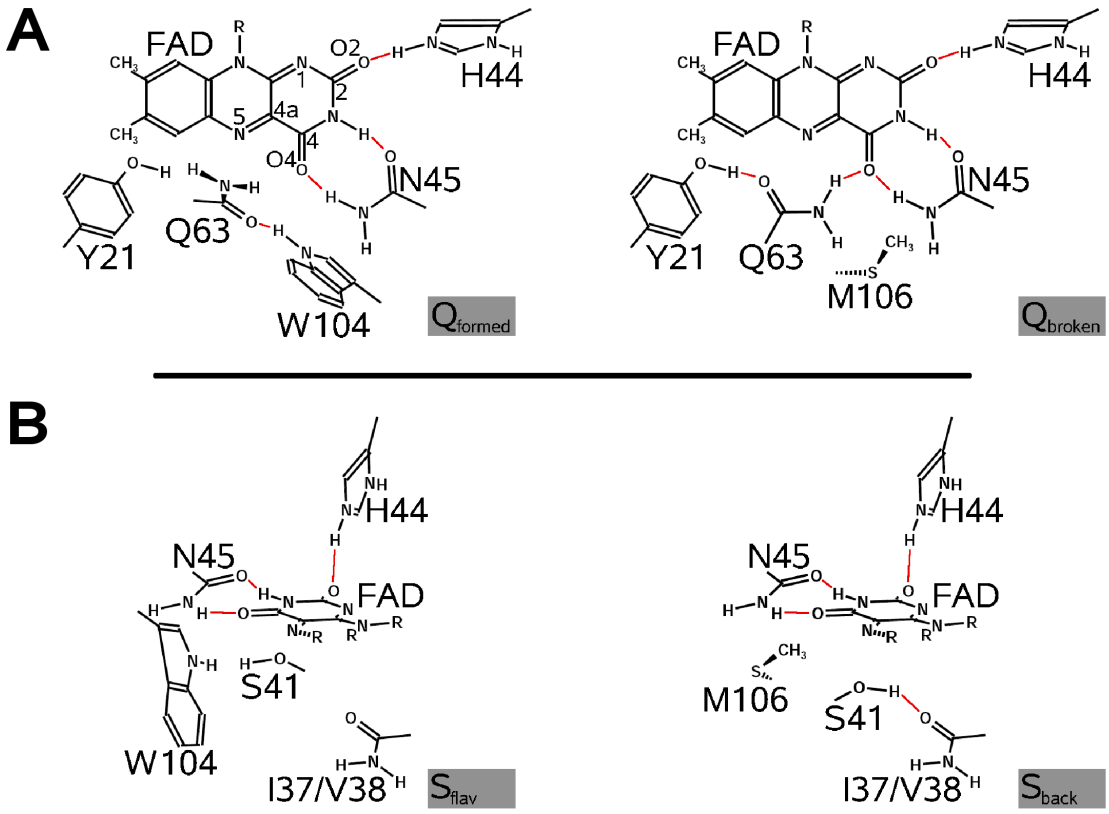

Figure 4.8: Schematic depiction of Gln63 (A) and Ser41 (B) positions, from [96]. Note that the presence and absence of Trp104 and Met106 is already a speculation on the possible consequences of the respective switch. It was not observed in the calculated structures, where Trp104 was always inside the BLUF domain.

Spectra of grouped and artificial structures

We omitted one of the NMR structures (#19), as it had Trp104 in a conformation not present in all other structures. Having 2 x 2 possible positions for the two amino acids Gln63 and Ser41, we sorted our remaining 19 NMR structures in four groups accordingly. The structural distribution in the NMR dataset was 13:6 for both Q_{formed}:Q_{broken} and S_{flav}:S_{back}. The 4 combined groups (Q + S) were distributed as follows: 8 structures were of the Q_{formed}/S_{flav} combination, Q_{formed}/S_{back} and Q_{broken}/S_{flav} had both 5 structures, and only 1 was of the Q_{broken}/S_{back} type. Averaging the spectra of the respective groups yielded the spectra in figure 4.9. However, the sampling size for each of the groups is now very small, especially for the Q_{broken}/S_{back} type. Thus, any qualitative conclusions from the fig. 4.9 have to be taken with care. MD structures were not analyzed as the sampling from the MD simulation could not be considered complete.

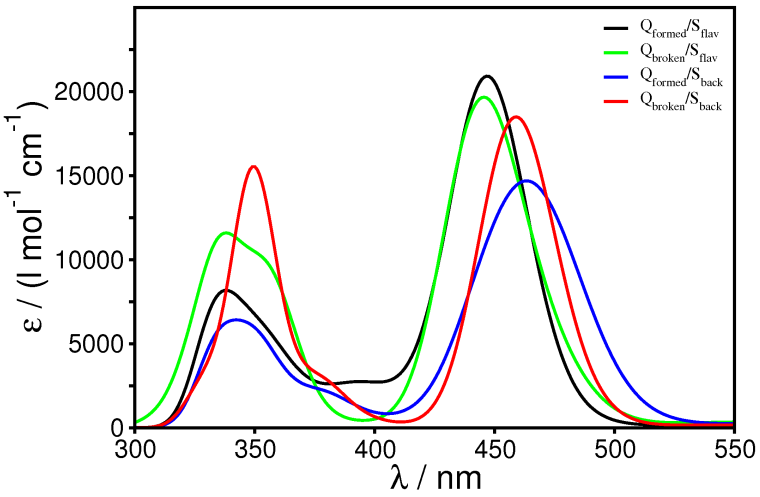

Figure 4.9: Average TD-B3LYP/6-31G(d) broadened spectra (30 singlet states) of the full model based on grouped NMR structures. No systematic behavior is observed. Groups indicated by legend. Broadening according to (4.1). Data partially published in [96].

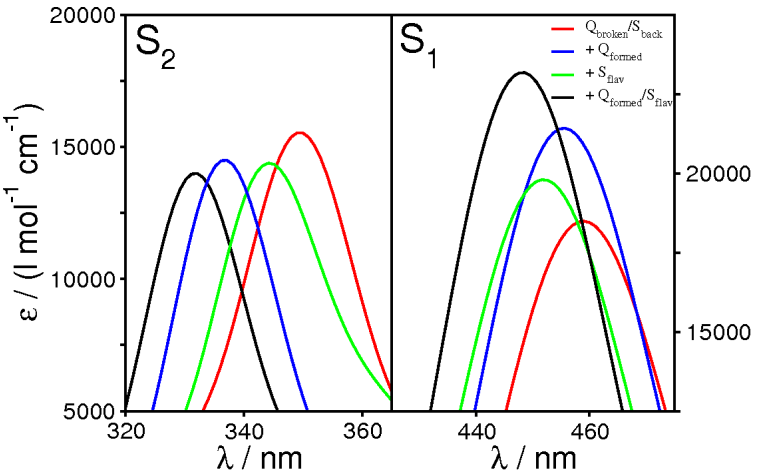

Figure 4.10: Stepwise evolution of the Q_{broken}/S_{back} geometry spectrum (red) when combined with the Q_{formed}/S_{flav} features of structure 2BUN#16. Artificial geometry features indicated by legend, broadening according to (4.1). Data partially published in [96].

As it is clear that the spectra are not influenced by Gln63 and Ser41 only, the missing of any systematic trend of the spectra in fig. 4.9 is understandable. It can only be seen that S_{back} is red shifted with respect to S_{flav} for $S_{0\to1}$, and that also the $Q_{broken} \to Q_{formed}$ change plays a role. In an attempt to make sure that these two amino acids are mainly responsible for the difference in the spectra, we artificially combined structural features of different geometries. As basis for these structures we took the single Q_{broken}/S_{back} conformation present in the 2BUN dataset (structure #4). We stepwise in-

troduced Q_{formed}/S_{flav} Gln63 and/or Ser41 coordinates relative to the N5, C4a and N10 atoms of the chromophore (for the flavin numbering, see fig. 4.1). We also had to transfer Tyr21 together with Gln63 or the Ile37/Val38 backbone segment together with Ser41 to avoid atoms overlapping each other. An exemplary result for these artificial structures is shown in fig. 4.10.
As can be seen, the structures having Q_{formed} and/or S_{flav} structures instead of the Q_{broken}/S_{back} conformations have distinct differences in the UV/vis spectrum. Both the $S_{0\to1}$ and $S_{0\to2}$ transitions of Q_{broken}/S_{back} are red shifted by approx. 10 nm compared to Q_{formed}/S_{flav}, which is in the order of the experimental shift upon formation of the signaling state. Interestingly, the intensity shifts are also similar to those found in the experiment [122, 92]. The trend shown in fig. 4.10 is qualitatively the same for all combination attempts performed with NMR structures #5, #9 and #16. The effect on the spectrum by changing either the Gln or Ser conformation is independent and cumulative: Both yield a small spectral shift to the red (about 5 nm each), with a slight change in band intensity.

From these investigations, we concluded that a change in the hydrogen bond network can indeed yield the observed UV/vis shifts observed in experiment. The participating hydrogen bonds may be interactions between FAD and the protein (Gln63 - FAD), but also purely inside the protein itself (Ser41 - backbone). The latter is problematic, as it increases the number of possible candidates for functional amino acids enormously.

4.3 Effects of Global BLUF Conformation

We need to see if the results from sec. 4.2 hold when they are transferred to other BLUF domains and conformations. Therefore we decided to employ a large-scale MD approach for several BLUF structures. This way, we aimed at a well-sampled set of structures for the proteins for which only X-ray or NMR structures are available.
In contrast to the previous section, we first created a series of snapshots from MD trajectories. This was followed by essentially the same investigations as were done in the last section, but extended by the presence of the protein environment in form of a point charge field.

4.3.1 Model Construction

We constructed three basic systems for each of the following PDB structures: 2BUN#9 (AppA-BLUF, Trp104 inside the domain), 2IYG (AppA-BLUF mutant, Trp104 outside the domain) and 2BYC (BlrB-BLUF[4], Trp104 outside the domain). This selection was based on the fact that one main feature of the signaling state is proposed to be a different Trp104 position [101, 102]. Which position is which state, however, is up to now not fully elucidated. A pictorial representation of the conformational differences in the in/out conformations of Trp104 (denoted W_{in} and W_{out} in the following) is shown in fig. 4.11. If our hypotheses from the previous section are correct, we would assume the Trp104 position to affect the hydrogen bonding network regarding Gln63 and Ser41 as well.

As a cysteine residue was replaced by serine in 2IYG, the structure was reverted to wild type AppA-BLUF by replacing the serine oxygen atom by sulfur. There were still minor sequential differences between the two AppA structures; please refer to the respective publications for further details [100, 104, 102]. The proteins were cut N-terminally before Asp15 and C-terminally after Leu108 (AppA structure and numbering; the analogous residues in BlrB are Glu and Ile).

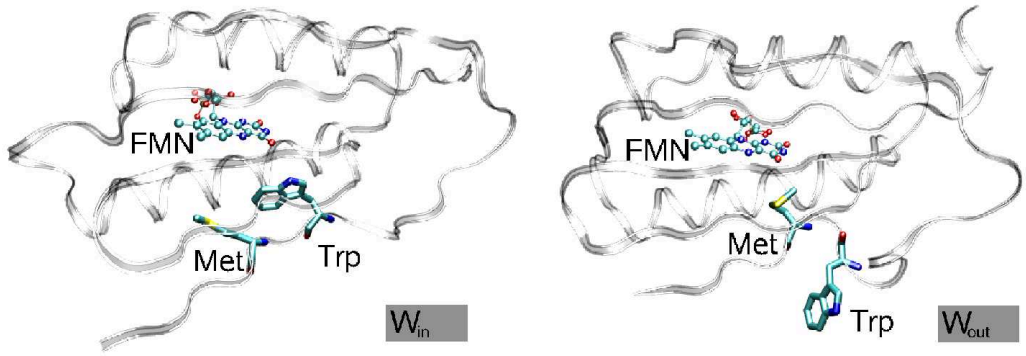

Figure 4.11: Comparison of BLUF W_{in} and W_{out} conformations. Left: AppA BLUF from PDB entry 2BUN#9; right: BlrB BLUF from PDB entry 2BYC. While most of the secondary structure is conserved, the Trp and Met residues compete for the position inside the protein close to the chromophore. Published in [96].

Most of the following calculations were carried out using the *Gromacs* package (v4.0.2) [123]. The trimmed protein constructs were solvated in a cubic box of TIP3P water [119]

[4]BlrB stands for "blue light receptor B", and is from the same organism as AppA, *Rhodobacter sphaeroides*.

using the Amber99 force field [53, 51] together with FMN as chromophore (parameters taken from [124]). If required, Cl^- ions were added to establish a neutral total charge of the box. Our criterion for the box size was a minimum distance from the solute to the box edges of 1 nm, resulting in different box vectors and number of water molecules for 2BUN#9, 2IYC and 2BYC: 5.6525 nm (5501 water molecules), 5.6525 nm (5459) and 6.2110 nm (7434), respectively. To improve comparability, we extended the slightly smaller box vector of the 2IYG system to the same length as for 2BUN#9, hence the equal value.

4.3.2 MD Protocol

Each of the three systems was propagated 8 times with a different random seed for 1.2 ns in a NVT setup (with T = 300 K) with periodic boundary conditions and a time step of 1 fs. We also tested 2 fs and 0.5 fs time steps, and found that 2 fs produced strongly different hydrogen bonding probabilities (see below, sec. 4.3.3), while the 0.5 fs yielded the same probabilities as 1 fs (maximum difference in a test trajectory below 5%, data not shown). The Coulomb long-range interactions were calculated using the Particle Mesh Ewald (PME) scheme [125], the cut-off distance was 1.2 nm for short-range interactions. Temperature coupling was performed using the v-rescale thermostat [58].

The resulting trajectories were checked visually for protein/chromophore dissociation or protein unfolding, which did not occur. The first 200 ps of the 1.2 ns propagation time were omitted as extended equilibration time (test trajectories had shown beforehand that equilibration was reached between 50 and 100 ps, data not shown). The remaining 1 ns trajectories (eight for each of the three models) were used as structural basis for the following analysis.

We now needed to gain a first insight on the effects of the BLUF species and their conformations on the hydrogen bonding of Gln63 and Ser41. This was done by analyzing the hydrogen bonds in the different geometries.

4.3.3 Hydrogen Bonding

Having seen that hydrogen bonding plays a decisive role for spectral details, we conducted a hydrogen bond analysis similar to what had been done before elsewhere for a smaller QM model [126]. To decide whether a hydrogen bond between an acceptor and a donor is present or not, we used the same criteria as in the previous work [126]: the donor-acceptor distance should be lower or equal to 3.1 Å (2.9 Å if both being oxygen), and the donor-hydrogen-acceptor angle is at least 120 degrees.

We applied this criterion to every 100th time step of a trajectory, yielding 10000 frames each. The fraction of frames with a hydrogen bond with respect to the total number of frames analyzed then produced the hydrogen bonding pattern for the respective trajectory (more than 10000 frames did not result in a change of hydrogen bonding patterns, data not shown). The hydrogen bonding patterns for each hydrogen bond were then averaged over the 8 trajectories for each protein conformation, resulting in the statistics given in table 4.3. The hydrogen bonds analyzed were selected based on their proposed importance for the photocycle found in several sources: The first row in table 4.3 directly arises from our findings in sec. 4.2. The two Gln63/FMN bonds were proposed elsewhere and are widely acknowledged as possible candidates for modulation of the UV/vis spectrum [92], while the Asn/FMN interactions were proposed in a recent publication [94].

Donor Acceptor	Ser41(OH) Ile37(O)	Trp104(NH) Gln63(O)	Tyr21(OH) Gln63(O)	Gln63(NH) Tyr21(OH)
AppA (W_{in})	0.412±0.358	0.000±0.000	0.829±0.086	0.000±0.000
AppA (W_{out})	0.954±0.018	0.000±0.000	0.877±0.062	0.000±0.000
BlrB (W_{out})	0.933±0.018	0.000±0.000	0.916±0.016	0.000±0.000
Donor Acceptor	Gln63(NH) FMN(O4)	Gln63(NH) FMN(N5)	FMN(N3H) Asn45(O)	Asn45(NH) FMN(O4)
AppA (W_{in})	0.184±0.148	0.021±0.026	0.194±0.169	0.330±0.086
AppA (W_{out})	0.015±0.007	0.036±0.009	0.143±0.029	0.428±0.078
BlrB (W_{out})	0.068±0.038	0.024±0.013	0.172±0.039	0.683±0.036

Table 4.3: Hydrogen bonding patterns of three different BLUF domain proteins/conformations, averaged over 8 trajectories each. Patterns are ratio of hydrogen bonded frames against total analyzed frames, listed with the respective mean deviation. Bonding patterns based on 10000 frames over 1 ns, see text for details. Upper table: Bonds between protein structures; Lower table: Bonds between protein and chromophore.

Although we observe small differences throughout the hydrogen bonding patterns of the three proteins/conformations, only a single hydrogen bond is strongly affected: Ser41 binds exclusively to the Ile37/Val38 backbone structure in the W_{out} conformations. The mean deviation, which is a measure of how ofter the hydrogen bond is formed or broken, is small in the case of the W_{out} Ser41-backbone hydrogen bond.
In the W_{in} conformation, Ser41 can be positioned differently (*i.e.*, the aforementioned S_{flav} position, see sec. 4.2.3), and in addition with a very large mean deviation: This means that the Ser41 residue is very flexible and can flip back and forth within a single trajectory between the two conformations S_{flav} and S_{back}. These switching events were studied further by extended trajectories (see below).
The other hydrogen bonds were found to be unaffected by the Trp conformation. While there might be a small influence on the Gln63-FMN(O4) interaction, the small absolute values forbid any clear statement based on our data.

Extended Trajectories

To study the lifetime and the occurrence of the assumed Ser41 switch, extended trajectories were calculated. The trajectories were set up analogous to the given MD setup above (sec. 4.3.2) with a total runtime of 10.2 ns. The first 200 ps were again omitted as equilibration time. The development of the Ser41 rotational angles is exemplary given in fig. 4.12 on the basis of such an extended 10 ns AppA (W_{in}) trajectory.

We set up eight 10 ns trajectories for each of the three proteins/conformations. The results did not allow for an estimation of a lifetime for the Ser41 positions, which seem to be at least in the order of several ns or more: Within the 10 ns period shown in fig. 4.12, a single switch of the Ser41 position occurs. From the exemplary trajectory given in fig. 4.12, we can estimate the switching event to complete within the range of 1 ps, as can be seen in the inset. For the W_{out} geometries, no Ser41 switching events were observed throughout all 10 ns trajectories, regardless of protein and in agreement with the small mean deviations for S41/I37 in table 4.3.

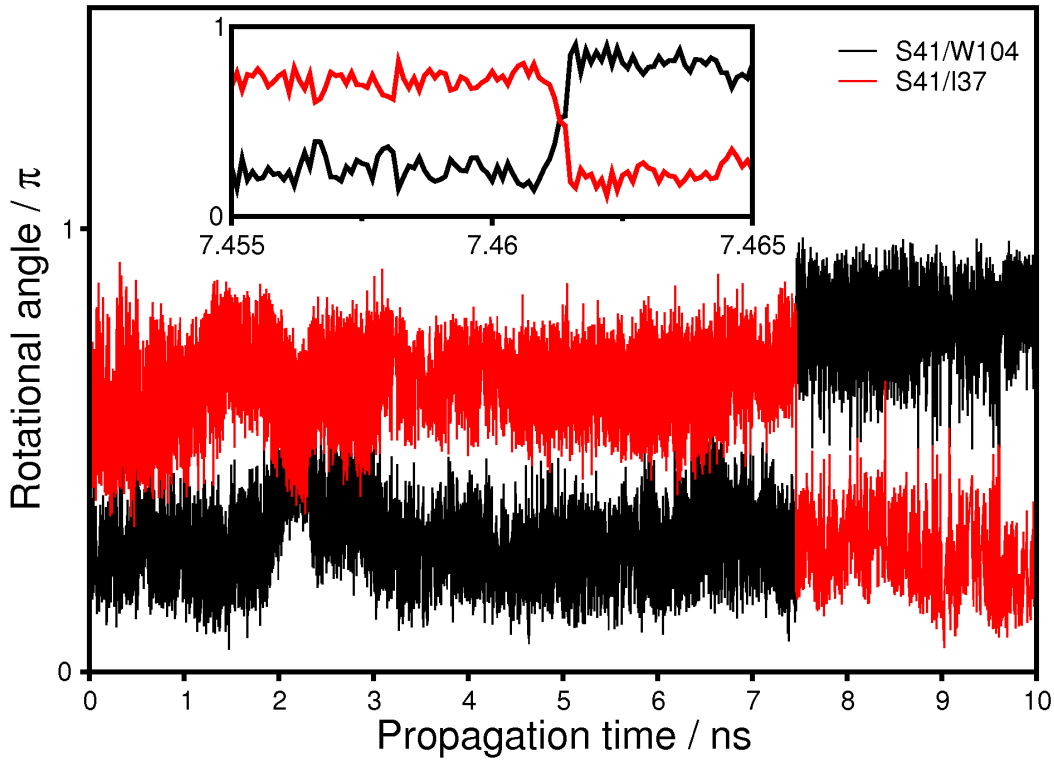

Figure 4.12: Rotational angle evolution of Ser41 within a 10 ns trajectory of AppA-BLUF (W_{in}). Inset: detailed view of switching event. Black: Ser41(HO-C_β-C_α) to Trp104(N_ϵ) dihedral angle; red: Ser41(HO-C_β-C_α) to Ile37(O) dihedral angle. Angle distributions correspond to S_{flav} before and S_{back} after the switching event.

4.3.4 MD Snapshot based UV/vis spectra

Analogous to our initial spectral characterization of the AppA BLUF NMR structures [100] (sec. 4.2), we aimed to find systematic dependencies of the UV/vis spectrum of the chromophore on the conformations. To calculate the averaged TD-DFT spectra for each of the proteins/conformations, we took for the QM part the model from sec. 4.2.1 with some modifications: We omitted the Met76 residue, as it did neither contribute directly to the S_1 HOMO, nor to its environment. Further, His44 had to be replaced in BlrB by arginine, due to a sequence exchange.

The global protein conformation can be expected to have effects surpassing the region close to the chromophore. Thus, we decided to include the omitted protein atoms and a 10 Å shell of water molecules around the protein as a field of point charges (electronic embedding). The charges of the atoms replaced by link atoms at the system border were omitted. The modified full model and the respective point charge field was extracted for

each of the three proteins/conformations for four of the trajectories: From each of the four 1.2 ns trajectories, we took the frames at 400 ps, 600 ps, 800 ps, 1000 ps and 1200 ps. Thus, we had 20 geometries with a corresponding point charge field for each of the three proteins/conformations, analogous to our initial characterization based on the NMR geometries.

Methodological difficulties (namely SCF convergence errors and small HOMO-LUMO gaps in the TD-DFT calculations) forced us to perform a QM/MM reoptimization of the structures for each MD snapshot[5]: The model as described above was taken as QM part, while the atoms of the point charge field were used as MM part. To couple the TURBOMOLE QM package to Amber [13], the ComQum interface was used [61]. Consequently, all QM calculations in the rest of this section were performed using TURBOMOLE 6.0 [62]. The optimization of the QM part was done with the BP86 functional and the 6-31G(d) basis set. BP86 is known to reproduce B3LYP geometries very well and is faster than B3LYP, as it is a pure density functional. During reoptimization, the MM part was frozen and treated with AMBER99 [51].

Based on the re-optimized snapshot geometries, we calculated TD-B3LYP/6-31G(d) (30 singlet states)[6] spectra in the point charge field. Out of the 20 structures, two for each protein/conformation did still experience unresolvable SCF problems or small HOMO-LUMO gaps, leaving us with 18 spectra for each system. The line spectra were broadened (using $\sigma = 750$ cm^{-1}) and averaged with the other spectra of their respective group. The resulting average spectra can be seen in fig. 4.13.

It is clearly seen that the approach used for simulating the spectra in fig. 4.13 is worse in terms of reproducing the experimental spectra than the simple approaches used in sec. 4.2, compare tables 4.1 and 4.2. Note that the experimental UV/vis spectra of all BLUF

[5]The reasons for the instabilities are up to now not fully explained. Test calculations using the Charmm force field instead of AMBER99 show no such problems. We have yet to see if the instabilities arose from the choice of force field or from the quantum chemistry package used. Other than that, there were no differences to the initial characterization of the BLUF domain, where such problems were much less frequent.

[6]Note that in the TURBOMOLE package, one has to select the "B3LYP-Gaussian" method to have the same correlation functional as in Gaussian09.

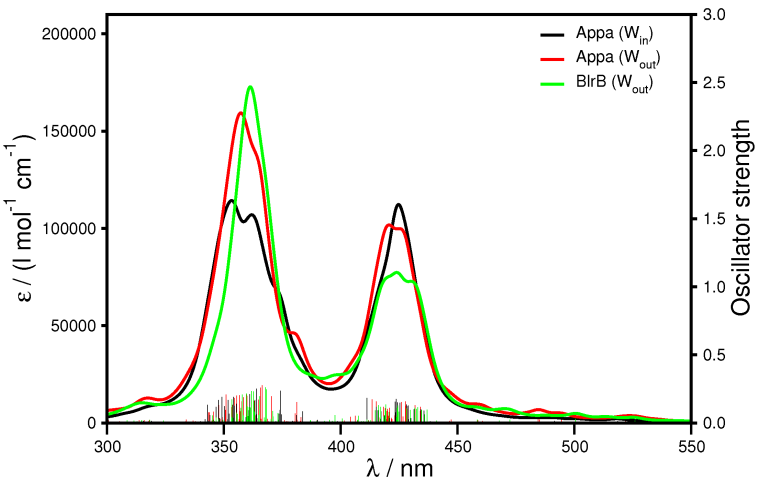

Figure 4.13: Averaged TD-B3LYP/6-31G(d) (30 singlet states) UV/vis spectra of the modified full model in a point charge field, based on 18 BP86/6-31G(d):AMBER99 re-optimized MD snapshots for each of the indicated proteins/conformations. See text for details. Broadened according to (4.1)

domains mentioned here are virtually identical [112, 92, 102]. Interestingly, while the intensity shifts are qualitatively as we would have expected if we assign W_{in} to the dark and W_{out} to the signaling state, the $S_{0\to2}$ transition is energetically not better located than in fig. 4.6, and the $S_{0\to1}$ maximum is shifted to the blue by about 25 nm.

If the Trp104 conformation was the main difference between the resting and the signaling state, we had expected the W_{out} proteins to be slightly different in terms of the UV/vis maxima from the W_{in} conformation, while the two W_{out} proteins were similar. However, fig. 4.13 shows no systematic difference between the two Trp104 conformations, W_{in} or W_{out}.

As we have seen that the Ser41 hydrogen bonding indeed changes and Ser41 has an influence on the spectrum, the overall similarity for all three proteins as seen in fig. 4.13 is puzzling. Test calculations (not shown) implementing a point charge field in the NMR structure approach of sec. 4.2 indicate that the point charge field is most probably not the reason, as spectra were still distinct. We may think of two reasons for the equivalent spectral behavior: First, the MM propagation does not yield a proper set of structures (18 snapshots may be too small, or important interactions like induced dipoles are missing in the MM approach). Second, one may consider the QM/MM optimization to eliminate the temperature based dynamics of the QM part, which is lost due to the reoptimization.

4.4 Summary and Discussion (BLUF)

We have shown that a small vacuum model is enough to reproduce of the experimental BLUF spectra, at least for the energetically lowest maximum. Our model was able to reproduce this $S_{0\rightarrow1}$ transition regardless of the structural basis (NMR or MD geometries). We found indications for certain amino acids (Gln63 and Ser41) to systematically influence the BLUF UV/vis spectrum, depending on their position and hydrogen bond partners. The latter of the two amino acids was up to recently experimentally mostly uncharacterized; but a late publication [127] confirmed that a S41A mutant of Slr1694 BLUF indeed seems to have a simpler resting state geometry (as it lacks a kinetic component in the excitation kinetics). Both resting and signaling state are red shifted in both $S_{0\rightarrow1}$ and $S_{0\rightarrow2}$ UV/vis transitions, too. Based on our model, we would exactly expect these two findings, as the flexibility of Ser41 introduces more possible geometries for the resting state (thus more kinetic components in the excitation kinetics). The Ser41 position should affect the spectrum as well, as we have shown in fig. 4.10. Since we did not find any contributions of the amino acids to the LUMO, a red shift of the UV/vis transitions into the LUMO indicates a destabilization of the participating occupied orbitals. The coefficients at Ser41 in the S_{flav} conformation would allow for an orbital stabilization, as the electron density becomes more delocalized.

Our more recent work with MD simulations on various BLUF structures has shown that the Ser41 residue is flexible only for the W_{in} conformation. This would indicate W_{in} to be the resting state (and hint at W_{out} being the signaling state). Unfortunately, methodological obstacles prevented us from using the MD sampled structures directly for the UV/vis spectra generation. We can only speculate that this may have resulted into the overall strong similarity of the calculated spectra in fig. 4.13.

A possible functionality for the Ser41 residue would be a kind of second anchor to fix Trp104, supporting the Gln63. We were unable to show any hydrogen bonding for Trp104-Gln63 in our simulations. However, this might be an artifact, as various sources indicate that there has to be such a hydrogen bond [101, 128]. A previous theoretical study [126] found this hydrogen bond to be even weakened in the W_{in} conformation,

while our results suggest otherwise. The main difference to this previous study is the use of a smaller time step in our simulations, which we found to affect hydrogen bonding patterns (data not shown).

If both Gln63 and Ser41 are anchors for Trp104 and W_{out} is the signaling state, it would be very straight forward to establish a possible series of events after photo-excitation, which is depicted in fig. 4.14. As we did not study the electron transfer in the system, we cannot speculate on how the events are triggered or even if the order of events is correct. Yet, low-temperature studies [129, 130] support our assumption of stepwise, spectrally important rearrangements during the formation of the signaling state: Accordingly, it seems possible that at low temperatures Ser41 and Gln63 switch independently or unconcerted, causing a small red shift each. The switching of Trp104 from inside to outside is functionally important, but has no effect on the UV/vis spectrum.

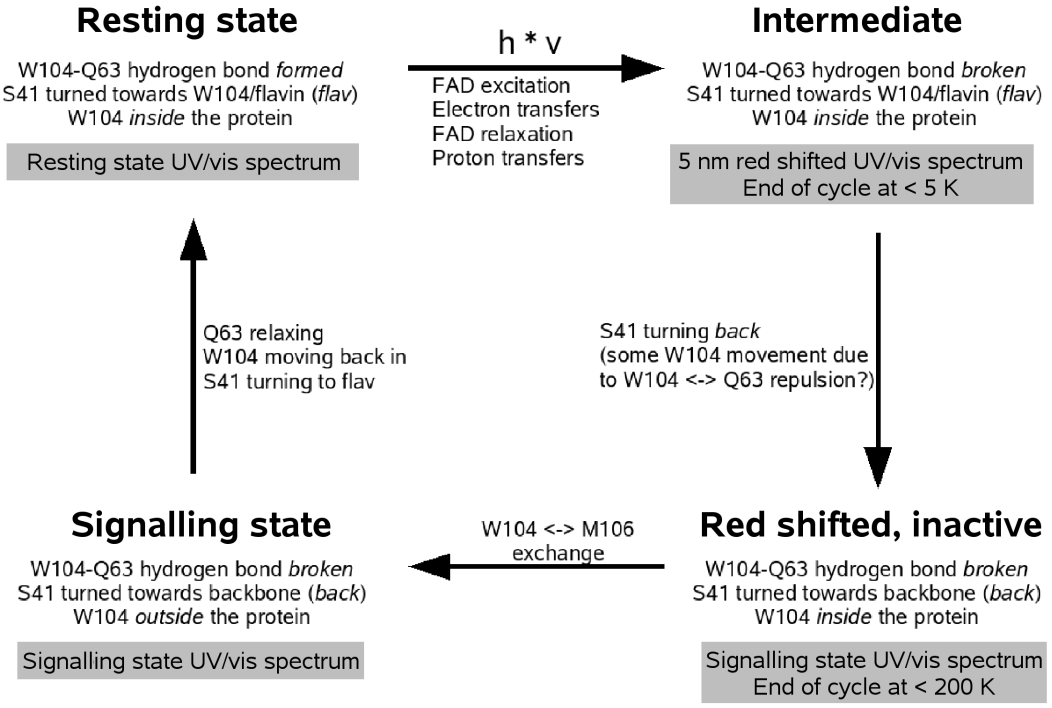

Figure 4.14: Series of proposed conformational events leading to the formation and relaxation of the AppA-BLUF signaling state. Temperatures refer to low temperature experiments stopping the photocycle at different positions at different temperatures [129, 130]. Adapted from [96].

Correlated *ab initio* studies of indole and water complexes (which can be seen as simplified versions of tryptophane and serine) have reported that there is indeed a confor-

mational minimum similar to the S_{flav} position with an interaction energy of 9.41 kJ/mol ($\approx$ 0.1 eV) [131]. This interaction is missing in all our force field approaches, but incorporation of this interaction would most likely support our model, as the existence of the S_{flav} position is of utmost importance for our findings.

As a direct consequence, the question arises how the electrons in the isoalloxazine ring react towards the permanent dipoles close to a position directly "above" the non-polar part (Tyr21) or "below" the polar part (Ser41). Exact correlation methods may be the key in extracting the necessary information about the true BLUF resting geometries: The common view of flavin domains, that all important interactions can be found close to the isoalloxazine plane, simply extending the π-system, may prove to be too limited.

Chapter 5

The LHCII Complex

5.1 The LHCII Energy Transfer

The Light Harvesting Complex II (LHCII) of chloroplasts transfers photonic energy to the photosynthetic reaction centers *via* excitonic coupling. LHCII contains both carotenoids and chlorophylls, which are the photoactive compounds. Chlorophyll is capable of directly accepting and transferring the energy of red light (about 650 nm). The carotenoids can support this light harvesting process by accepting photons of different wavelengths and transferring this energy to the chlorophyll molecules [16].
Carotenoids can also serve as protective units, preventing the excited electrons from damaging the complex [16, 132]. Thus, a carotenoid can either function as energy donor or acceptor for the first excited state of chlorophyll. In this chapter, we have characterized the interaction of one such donor and chlorophyll.

Several sources have reported studies regarding the nature and predictability of the participating excited states [133, 134, 135, 136, 137, 138]. The influence of the protein environment, however, is up to now almost unstudied. The aim of this chapter is therefore the construction of an approximate "LHCII plus chromophores" model system. We will examine the effects of an approximated environment on the UV/vis transition processes in the chromophores predicted by various theoretical methods.

5.1.1 Structure and Function of LHCII*b*

LHCII*b* is a nuclear encoded protein family with several members. LHCII*b* proteins are known to form membrane embedded trimer complexes, binding both chlorophyll and carotenoids. Their purpose is to spatially arrange both the chlorophyll and carotenoid molecules. This enables exciton transfer to the core antenna complexes which themselves are excitonically coupled to the photosystems I and II. This way, the thylakoid membrane inside the chloroplasts is more efficient in acquiring the energy, which is needed by the photosystems to keep the light reactions of photosynthesis running [69].

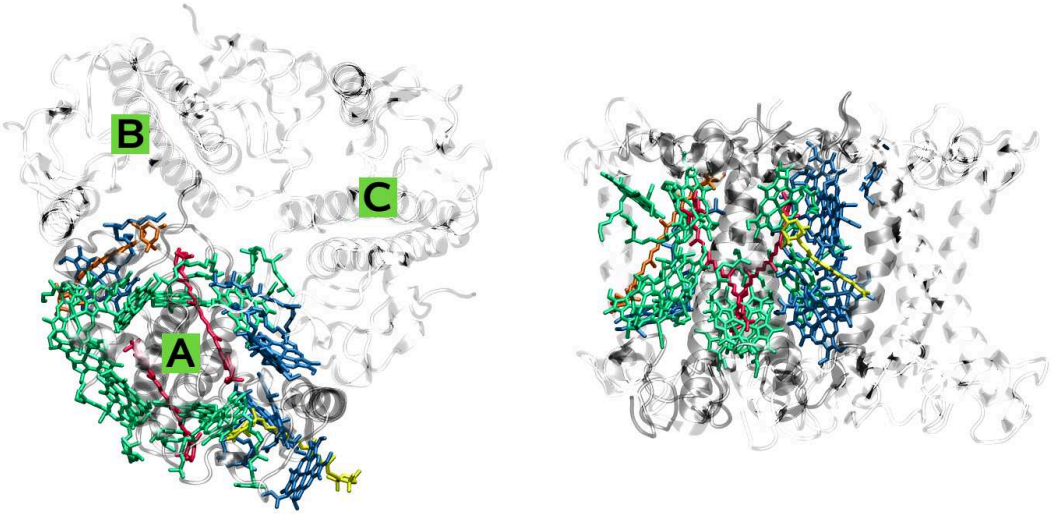

Figure 5.1: Display of the pea LHCII*b* trimer complex as in PDB dataset 2BHW [132]. Left: "Top" view, as if looking on the thylakoid membrane plane from the luminal side. Right: "Side" view, as if looking at a thylakoid membrane cross section. Monomer A highlighted, monomers B and C shown schematically. Green: chlorophyll a; blue: chlorophyll b; red: lutein; yellow: 9'-cis-neoxanthin; orange: violaxanthin.

The overall geometry of the LHCII of *Pisum sativum* (pea) is depicted in fig. 5.1 [132]. We will use the corresponding numbering and geometry for further presentation of the LHCII features. It can be seen that every monomer features several non-protein structures: Two luteins are spanning roughly from one membrane side to the other. Close to one monomer/monomer interface, a 9'-cis-neoxanthin is located, while on the other interface, a violaxanthin is present. The binding site of violaxanthin may also be occupied by zeaxanthin [16].

Most dominating throughout the structure, however, is the presence of chlorophyll: We find 8 chlorophyll a and 6 chlorophyll b molecules arranged throughout the whole structure

of a single monomer. Additionally, we find membrane glycerols in the protein structure. This is not surprising as the protein spans through the thylakoid membrane from the lumen to the stroma: It requires the presence of these molecules to establish its native conformation in the crystal structure.

The various light absorbing molecules in the LHCII enable the thylakoid membrane to harvest more light from the visible spectrum as would be possible with the photosystems I and II alone. The energy collected by the various chlorophyll molecules in LHCII is transferred to the photosystems *via* excitonic coupling. Several carotenoids are located closely to the chlorophyll molecules. Thus, we can expect the carotenoids and chlorophylls to interact, most probably also by excitonic coupling. Yet, the exact purpose of the coupling is determined by the type of the carotenoid, and can be either supporting the population of the chlorophyll S_1 state (*light harvesting*) or act as a quencher of chlorophyll fluorescence (*non-photochemical quenching*, NPQ). The role of violaxanthin, a carotenoid suspected to help in light harvesting, will be discussed now.

5.1.2 Light Harvesting

Violaxanthin is a carotenoid suspected to play a role for the light harvesting process [139]. It is interchanged *via* the violaxanthin cycle with zeaxanthin, which is supposed to be its counterpart in excess light, performing non-photochemical quenching (NPQ) [140, 141]. The structure of violaxanthin in shown in fig. 5.2.

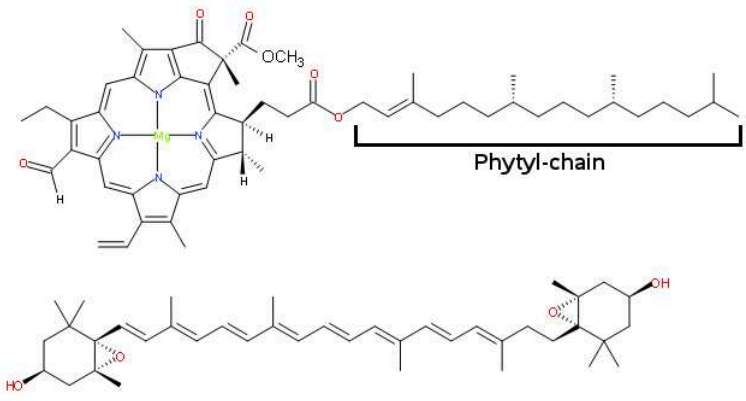

Figure 5.2: The structures of chlorophyll b (top) and violaxanthin (bottom).

Experimental studies show that the order of the excited state energies connected to violaxanthin in solution allows for a certain scheme of the light harvesting mechanism [16, 142], which is given in fig. 5.3. As the ground and the first excited state of violaxanthin are both of ${}^1A_g$ symmetry in an idealized C_{2h} model, the $S_{0\to1}$ excitation is optically forbidden.[1] Since the S_1 is required for the harvesting mechanism, the energy from the strong one-electron HOMO $\to$ LUMO transition ($S_{0\to2}$, S_2 being of ${}^1B_u$ symmetry) has to dissipate first, relaxing violaxanthin to the S_1 state before energy transfer to chlorophyll. The two S_1 states of chlorophyll and violaxanthin are then supposed to be excitonically coupled, allowing for a direct energy transfer.

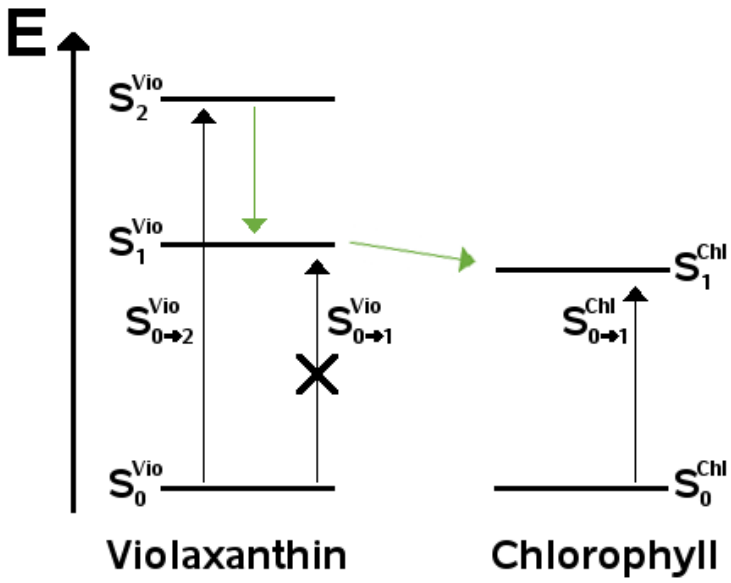

Figure 5.3: Violaxanthin/chlorophyll energy levels of the free molecules in solution, schematically. The energetic order enables violaxanthin to transfer energy to chlorophyll, populating the chlorophyll S_1. However, direct excitation of the violaxanthin ($S^{Vio}_{0\to1}$) is forbidden due to symmetry reasons. Thus, the S_2 has to be populated first before energy transfer (green arrows).

Chlorophylls of the LHCII are usually excitonically coupled to transfer the energy acquired in the S_1 state to the core antennae and finally to the photosystem II [69, 143]. However, some downstream reactions are diffusion controlled and thus several orders of magnitude slower than the exciton transfer processes. It is thus likely to happen under light stress that the downstream energy acceptors are saturated and incapable of taking the energy from the antennae/LHCs. As such, a conversion from a light harvesting enhancing mechanism to the NPQ photoprotection prevents the photochemical degradation

[1] It is found that this approximation reflects the optical properties of carotenoids very well [16], even though they formally exhibit C_2 symmetry in most cases.

makes sense in excess light. This is supported by the fact that violaxanthin is converted to zeaxanthin upon lumen acidification [144].

We will discuss the process of initial light absorption in the light harvesting configuration of LHCII in the rest of this chapter. This is due to the fact that the available crystal structure of pea LHCII [132] contains violaxanthin, and not zeaxanthin.

5.2 Modeling of Initial Light Absorption in Pea LHCII

5.2.1 Model Construction

Our models of the violaxanthin/chlorophyll b (Vio/Chl) pair was based on pea (*Pisum sativum*) LHCII, PDB dataset 2BHW [132]. We chose the chromophore pair from the chain A of the LHCII trimer in 2BHW. An exemplary model system can be seen in fig. 5.4, the location of the chromophore pair in the complex can be seen in fig. 5.1. We also recalculated the most important spectra again for monomers B and C, the results can be seen in appendix A.2.1.
Any geometry optimizations in this section were performed with B3LYP/6-31G(d) using Gaussian09 [40]. Hydrogens were added manually to the crystal structure geometries of Vio and Chl, and optimized *in vacuo* leaving the other atoms frozen. From these structures, we devised a series of models:

- 1^{Vio} and 1^{Chl}: Fully optimized, isolated Vio or Chl
- 2^{Vio} and 2^{Chl}: Isolated Vio or Chl *in vacuo* in crystal geometry (with *in vacuo* optimized hydrogens)
- 3: Combined Vio and Chl in crystal geometry (with *in vacuo* optimized hydrogens)

To analyze the excited states in terms of symmetry, we also calculated the spectra of a Vio in enforced C_2 symmetry.
For comparison with experimental data, we calculated both models 1 *in vacuo* and in a polarized continuum model (PCM) with acetone as solvent [39]. Note that models 2^{Chl}

and 3 were supported by 3 amino acids to fill the Chl Mg^{2+} coordination sphere: Lys23, Tyr24 and Leu25. The amino acid chain was considered to be uncharged and in crystal geometry. The backbone was cut at the peptide bond, N-terminus was NH_2, C-terminus CO_2H, manually added atoms were optimized together with the hydrogens. The effect of the amino acids on the spectrum were however negligible, comparative spectra can be found in appendix A.2.2.
We performed similar test calculations for the phytyl residue (appendix A.2.3). It was found to affect the $S^{Vio}_{0\to2}$ and $S^{Chl}_{0\to1}$ transitions slightly, this will be discussed later. The phytyl residue was included in all calculations presented in this section.

Since the models 2^{Vio}, 2^{Chl} and 3 were directly cut from the crystal geometry, it was possible to embed these into an environment in form of a point charge field (PCF). The PCF was constructed as follows: We used the Amber99 force field [51] to obtain point charges for the protein monomers (missing heavy atoms were added manually, hydrogens were added using the Gromacs 4.0.2 program pdb2gmx [49, 53]). Glu, Asp, Lys and Arg were considered to be charged in all cases, His was considered to be δ-protonated only. This resulted in a total charge of -10 per monomer. To see if this strong overall charge has an effect on our results, we calculated selected spectra with fully protonated histidines, too (see appendix A.2.4).
For all non-protein molecules in the PDB dataset, we used the crystal geometry, added hydrogens manually and optimized them using PM6 [145]. For these conformations, we calculated the point charges using B3LYP/6-31G(d,p) and the electrostatic potential (ESP) scheme [42, 43]. To make sure that the point charges closely resemble the charges found in a protein environment, we embedded the non-protein molecules in a PCM using chloroform as solvent for all corresponding calculations. It was previously proposed that these conditions are similar to these found inside a protein [82].
Having obtained all point charges for all atoms in the PDB dataset, we were able to embed any molecule of interest in a PCF. This was done by simply omitting the point charges for the respective molecule and placing the explicitly treated molecule at the position of the omitted charges. This treatment yielded three further models, labeled 2^{Vio}+PCF, 2^{Chl}+PCF and 3+PCF, analogous to the model designation given above. The hydrogens

of the explicitly treated molecules were in the same positions as in the *in vacuo* models (no reoptimization upon PCF introduction).

For models 1, we calculated a TD-DFT spectrum of 30 singlet states, using B3LYP, BHandHLYP and CAM-B3LYP and the 6-31G(d) basis. It was found in test calculations (see appendix A.2.5), that there is no significant difference in the spectrum when using the 6-31G(d,p) basis, so we used the smaller basis to save computational resources. We also conducted calculations with the CIS method [33] within the same basis for comparison, which can be found in appendix (A.2.6).

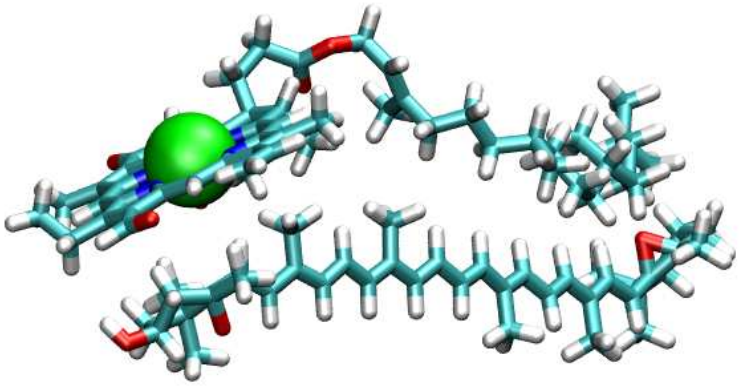

Figure 5.4: Exemplary model system (model 3+PCF), point charges not displayed. 236 atoms, 816 electrons (1906 basis functions in 6-31G(d) basis).

The calculated line spectra were broadened according to (4.1) with a σ of 750 cm^{-1}.

5.2.2 Isolated Molecules

General remarks

All calculated spectra presented here are vertical absorption spectra, which are expected to show absorption at wavelengths equal or lower in energy compared to the vibronic 0 $\rightarrow$ 0 transition [116]. Since both chromophores display a distinct vibronic fine structure of their UV/vis bands, our calculated transitions cannot reproduce the spectral maxima exactly. Thus, all results are qualitative in nature.

It has been shown elsewhere that it might be possible to determine excited states with strong double-excitation character (like the S_1^{Vio}) using TD-DFT [134, 135]. They are, however, badly reproduced and it is not advisable to rely on TD-DFT calculations for the

respective excitation energies. As such, we will discuss only the single-electron HOMO → LUMO transitions, which dominate the character of $S^{Vio}_{0\to2}$ and $S^{Chl}_{0\to1}$ [135].

Band assignment

To see if the given calculations reproduce the Vio state symmetries discussed above and elsewhere [16], we calculated the spectral features of fully optimized Vio in enforced C_2 symmetry with B3LYP/6-31G(d).

The calculated first excited state was of B symmetry and was dominated by the HOMO → LUMO transition. The respective oscillator strength was high (3.7436), and the band maximum was found to be at 522 nm (2.3773 eV).

We concluded that this transition corresponds to the $S^{Vio}_{0\to2}$ transition, as it is clearly optically allowed and of high HOMO → LUMO one-electron character. It is known that the S_2 state of linear polyenes corresponds to a simple HOMO → LUMO transition [16, 135]. Our methods therefore deliver the S^{Vio}_2 state as seemingly first excited state, just as has been found before elsewhere [135]. We will assign the labels S^{Vio}_2 and $S^{Vio}_{0\to2}$ regardless, to be consistent with the experimental terminology [16, 142].

There is no higher symmetry than C_1 possible for Chl. Thus, we have to rely on our other calculations for the orbital assignment (see below). We will see, however, that the calculated spectra behave as proposed by experiment [146] and the participating orbitals are the Chl HOMO and LUMO. Accordingly, we have assigned the first calculated transition to be the experimentally observed $S^{Chl}_{0\to1}$, which is mainly of one-electron HOMO → LUMO character. Additionally, it is comparatively weaker than the higher Chl transitions (see fig. 5.6 for comparison). This was expected, since this can be derived already from simplified symmetry considerations given elsewhere [146].

Optimized systems

To benchmark our methods to the experimental findings in acetone [16], we compare the features for each state with strong HOMO → LUMO character with the experimental values (table 5.1). The spectra for CAM-B3LYP and B3LYP are shown exemplary in figs. 5.5 and 5.6.

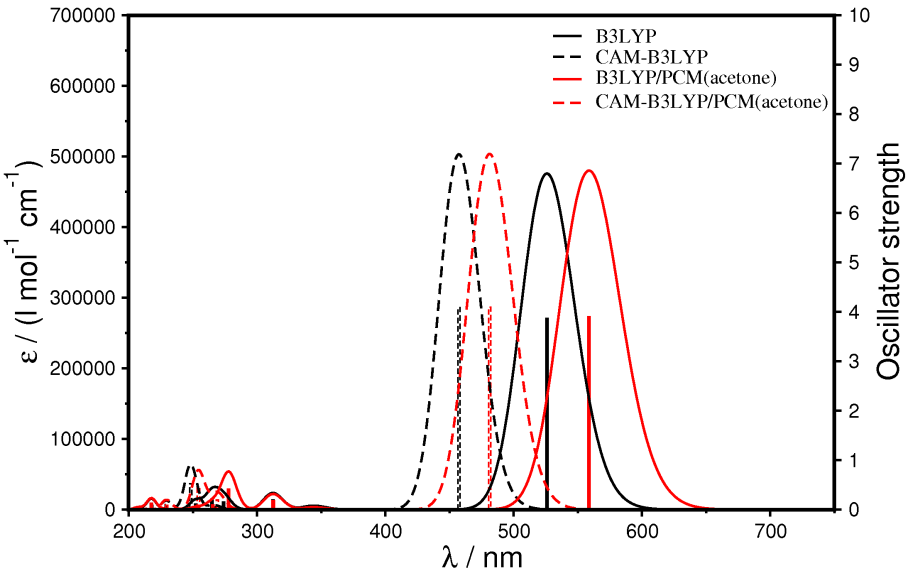

Figure 5.5: (CAM-)B3LYP/6-31G(d) spectra of Violaxanthin in fully relaxed B3LYP/6-31G(d) geometry (model 1^{Vio}), *in vacuo* (black) or in an acetone PCM (red).

It can be seen that $S^{\mathrm{Vio}}_{0\to2}$ is much more intense than $S^{\mathrm{Chl}}_{0\to1}$. It should be recognized that the $S^{\mathrm{Vio}}_{0\to2}$ maximum can come unrealistically close to the $S^{\mathrm{Chl}}_{0\to1}$ maximum in the B3LYP case. This will become important later on.

All DFT methods produce similar results for Chl in terms of $S^{\mathrm{Vio}}_{0\to2}$ and $S^{\mathrm{Chl}}_{0\to1}$ absorption maxima. CIS maxima are at too short wave lengths. Yet, the CAM-B3LYP values for Vio in acetone correspond better to the experimental values than the ones of all other methods. From the results of model 1 in all variations (table 5.1), we can already conclude that CAM-B3LYP is the best choice for reproducing the spectra of both isolated chromophores. This underlines the proposed strength of the Coulomb-attenuating method for the polarization of long chains [32]. We kept B3LYP regardless for comparative purposes, as it is a widely used standard method.

Isolated systems in crystal geometry

To see which effect the geometrical distortion within the crystal has on the chromophores, we will now compare the values for the optimized, *in vacuo* models 1^{Vio} and 1^{Chl} with the results for the crystal geometry, *in vacuo* models 2^{Vio} and 2^{Chl}. The details for models 2 are listed in table 5.1.

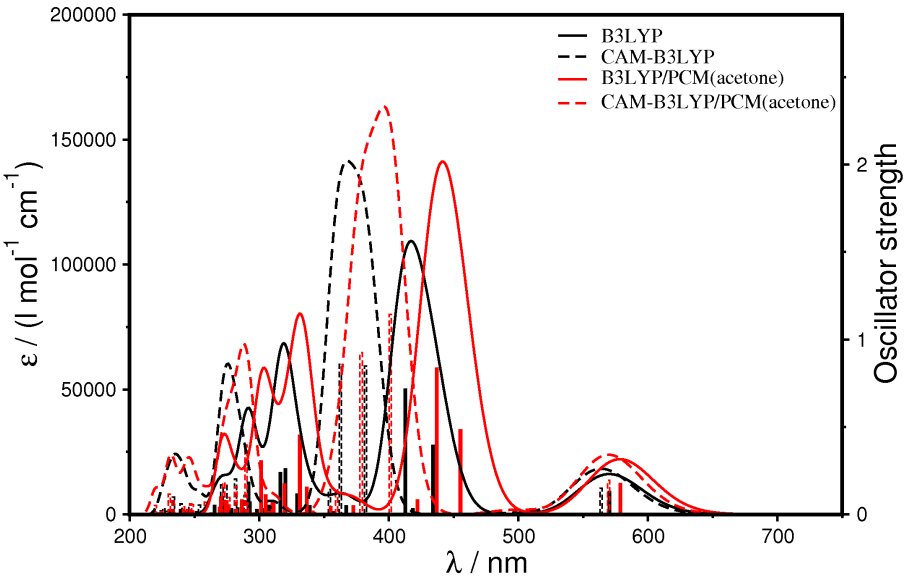

Figure 5.6: (CAM-)B3LYP/6-31G(d) spectra of Chlorophyll b in fully relaxed B3LYP/6-31G(d) geometry (model 1^{Chl}), *in vacuo* (black) or in an acetone PCM (red).

We see a red shift for both chromophores, which is much more pronounced for violaxanthin: The $S^{\mathrm{Vio}}_{0\to 2}$ shifts by 0.34 eV to lower energy for B3LYP. In the CAM-B3LYP case, the maximum shifts by 0.56 eV in the same direction, compared to model 1^{Vio}. The red shift of $S^{\mathrm{Chl}}_{0\to 1}$ is smaller, the maximum being at 600 nm for both B3LYP and CAM-B3LYP. The shift is therefore only 0.11 eV or 0.13 eV, respectively. The crystal structure geometry therefore has a much larger effect on the S^{Vio}_2 than on the S^{Chl}_1.
Note that the energetic order of the states is inverted compared to models 1 in the B3LYP case. With this order of states, light harvesting would be impossible or extremely hindered, as $S^{\mathrm{Vio}}_{0\to 2}$ is energetically lower than $S^{\mathrm{Chl}}_{0\to 1}$. We will see that this feature of the B3LYP spectra will remain the same throughout the rest of this work.

We conclude that the energetic shift is larger for Vio. For a full characterization of the isolated molecules, we will investigate the influence of the point charge field (PCF) on the molecules in crystal geometry in the following.

Isolated systems in PCF

Models 2^{Vio}+PCF and 2^{Chl}+PCF essentially surround the isolated chromophores in crystal geometry by a point charge field mimicking all other atoms of the PDB dataset,

		Chlorophyll b	Violaxanthin
Model	Method	λ_{max}/nm (eV)	λ_{max}/nm (eV)
1 *in vacuo*	B3LYP	570 (2.18)	526 (2.36)
	BHandHLYP	551 (2.25)	458 (2.70)
	CAM-B3LYP	564 (2.20)	457 (2.71)
	CIS	487 (2.55)	375 (3.31)
1+PCM (acetone)	B3LYP	578 (2.15)	559 (2.22)
	BHandHLYP	556 (2.23)	484 (2.56)
	CAM-B3LYP	570 (2.18)	482 (2.57)
	CIS	415 (2.99)	501 (2.48)
Exp. [147, 148]	(acetone)	645 (1.95)	470 (2.64)
2 (crystal geom., *in vacuo*)	B3LYP	600 (2.07)	615 (2.02)
	CAM-B3LYP	600 (2.07)	577 (2.15)
2+PCF (crystal geom., with PCF)	B3LYP	608 (2.04)	633 (1.96)
	CAM-B3LYP	599 (2.07)	593 (2.09)

Table 5.1: Overview of the isolated Chlorophyll b and Violaxanthin band maxima with strong HOMO → LUMO character (models 1 and 2). Basis set is always 6-31G(d).

including the hydrogens. The resulting TD-DFT maxima are listed in tab. 5.1. We see, again, that the $S^{Vio}_{0\to2}$ maximum is minimally red shifted for B3LYP and CAM-B3LYP in comparison to the spectra of models 2^{Vio}. $S^{Chl}_{0\to1}$ remains almost unaffected.
The red shift induced by the PCF is much smaller than the one arising from the crystal geometry. As before, the shift is larger for $S^{Vio}_{0\to2}$ than for $S^{Chl}_{0\to1}$, independent of the DFT functional. The energetic order of states is different for CAM-B3LYP and B3LYP, just as observed without the PCF.

We have now an overview of the properties of the isolated chromophores in different geometries and environments. For the investigation of the interaction of both chromophores we will look at both molecules together in the same model in the following section.

5.2.3 Coupled chromophore systems

We devised two models, 3 and 3+PCF, which essentially consist of both chromophores in crystal geometry (like models 2). 3+PCF surrounds model 3 with the appropriate PCF, analogous to models 2+PCF and 2. The spectra for models 3 and 3+PCF are shown in figs. 5.7 and 5.8, respectively. As the electronic structure is not as simple as for the

isolated molecules, we have to rely on a more sophisticated, detailed analysis for each observed transition in both models. These details can be found in table 5.2.

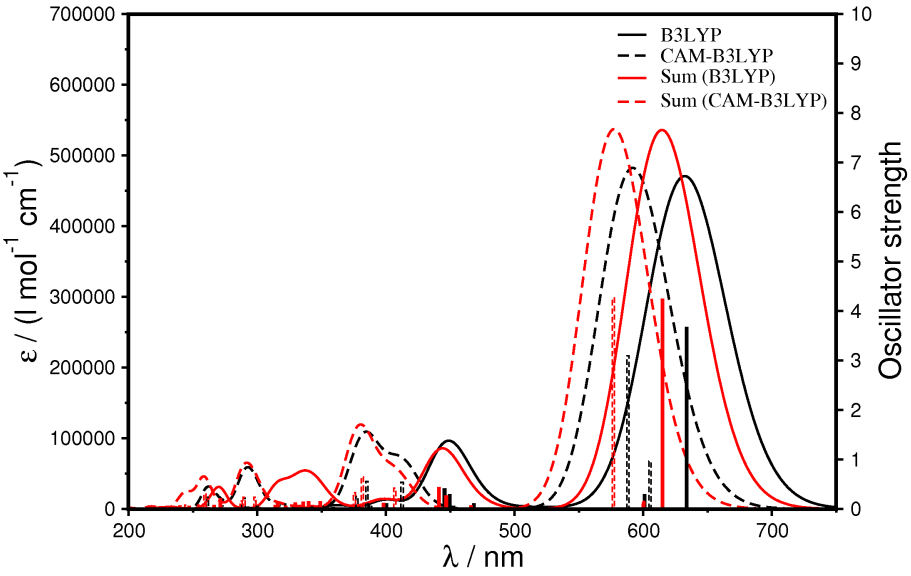

Figure 5.7: (CAM-)B3LYP/6-31G(d) spectra of the Violaxanthin/Chlorophyll b chromophore pair in crystal geometry (model 3), supported by a amino acid chain for Mg^{2+} coordination. See text for details. The red lines (sum) show the spectra resulting from the summation of the spectra for model 2^{Vio} and 2^{Chl}.

In the following, the assignment of the excited states is based on oscillator strengths. The spectra always showed two bands in the region of about 600 nm. The stronger band was assigned the label $S^{\mathrm{Vio}}_{0\to2}$, the weak band was assigned the label $S^{\mathrm{Chl}}_{0\to1}$, analogous to the single molecule spectra.

Consequently, the orbital assignment in table 5.2 is based on the coefficients for the respective electronic excitations: The orbitals participating in the $S^{\mathrm{Vio}}_{0\to2}$ transition with the highest coefficient[2] were assigned to be the Vio_{HOMO} and Vio_{LUMO}, although they correspond to the HOMO and the LUMO+1 for the system as a whole. The same criteria were applied to the $S^{\mathrm{Chl}}_{0\to1}$ for the Chl_{HOMO} and Chl_{LUMO}, which are the HOMO−1 and the LUMO for the whole system. A visualization of the respective orbitals is depicted in fig. 5.9. It can be seen that our band assignment based purely on the oscillator strength is supported by the localization of the participating orbitals: The orbital transitions of the coupled system correspond well to Chl or Vio HOMO → LUMO transitions, like in

[2]The sum of all (coefficients)^2 is normalized to 0.5 in the Gaussian output, as two electrons from each orbital can participate in each orbital transition. Hence, to get the participation of an orbital transition in total, the (coefficients)^2 must be simply doubled.

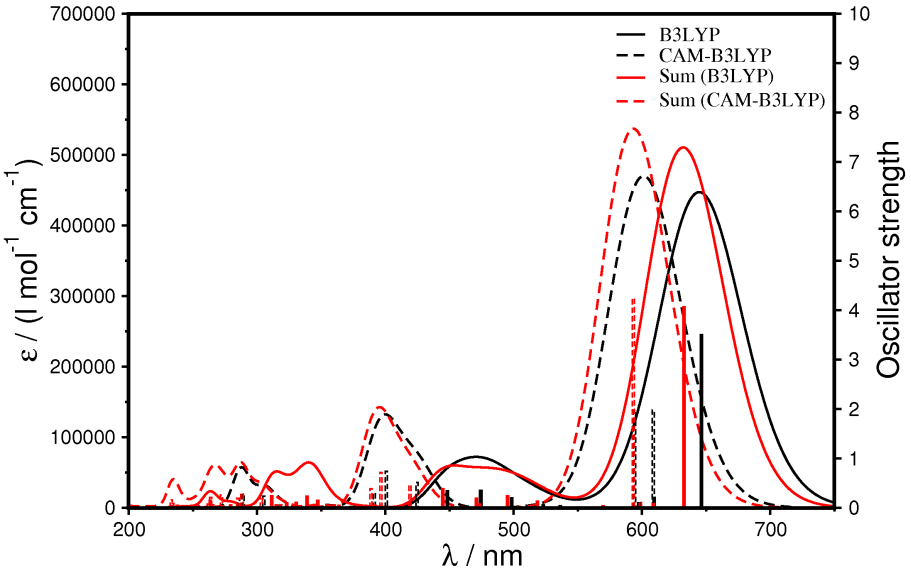

Figure 5.8: (CAM-)B3LYP/6-31G(d) spectra of the coupled Chlorophyll b/Violaxanthin chromophore pair in crystal geometry (model 3+PCF), surrounded by a point charge field. See text for details. The red lines (sum) show the spectra resulting from the summation of the spectra for model 2^{Vio}+PCF and 2^{Chl}+PCF.

the isolated models.

Models 3 and 3+PCF show again red shifted band maxima compared to models 2 and 2+PCF, respectively. The shift is small (approx. 0.05 eV), but observed for all cases. We have already shown that the distinction between $S^{\mathrm{Vio}}_{0\to2}$ and $S^{\mathrm{Chl}}_{0\to1}$ is more difficult, as HOMO $\to$ LUMO character from each individual chromophores is found in both excited states. Moreover, it is difficult to apply the criteria used above to the CAM-B3LYP case, where the oscillator strength is split evenly (about 1.97 for both $S^{\mathrm{Vio}}_{0\to2}$ and $S^{\mathrm{Chl}}_{0\to1}$). This type of electronic coupling between the two CAM-B3LYP excitations is also represented by the coefficients, which are of similar size for both HOMO $\to$ LUMO transitions in both excited states (see tab. 5.2).

Since we based above calculations only on one of the three available LHCII monomers, we also calculated the CAM-B3LYP results using the monomers B and C of the PDB dataset (see fig. 5.1). The aim was to make sure that the findings are not based on unique structural parameters in monomer A. It was found that the observed trends are reproduced for monomers B and C. However, they are not as pronounced as in the original monomer A (see appendix A.2.1).

Model	B3LYP/6-31G(d)				
	λ_{max}/nm (eV)	f_I	Character	Coefficient	%
3 *in vacuo*	634 (1.96)	3.6673	$Chl_{HOMO} \rightarrow Chl_{LUMO}$	0.14137	4.0%
			$Vio_{HOMO} \rightarrow Vio_{LUMO}$	0.52724	55.6%
	601 (2.06)	0.2889	$Chl_{HOMO} \rightarrow Chl_{LUMO}$	0.56599	64.1%
			$Vio_{HOMO} \rightarrow Vio_{LUMO}$	-0.10867	2.4%
3+PCF	646 (1.92)	3.5035	$Chl_{HOMO} \rightarrow Chl_{LUMO}$	0.13274	3.5%
			$Vio_{HOMO} \rightarrow Vio_{LUMO}$	0.70439	99.2%
			$Vio_{HOMO} \leftarrow Vio_{LUMO}$	-0.17698	-6.3%
	609 (2.04)	0.2140	$Chl_{HOMO} \rightarrow Chl_{LUMO}$	0.46417	43.1%
			$Vio_{HOMO} \rightarrow Vio_{LUMO}$	-0.11836	2.8%
Model	CAM-B3LYP/6-31G(d)				
	λ_{max}/nm (eV)	f_I	Character	Coefficient	%
3 *in vacuo*	605 (2.05)	0.9630	$Chl_{HOMO} \rightarrow Chl_{LUMO}$	0.56191	63.1%
			$Vio_{HOMO} \rightarrow Vio_{LUMO}$	0.32910	21.7%
	588 (2.11)	3.0949	$Chl_{HOMO} \rightarrow Chl_{LUMO}$	-0.28565	16.3%
			$Vio_{HOMO} \rightarrow Vio_{LUMO}$	0.61484	75.6%
3+PCF	609 (2.04)	1.9736	$Chl_{HOMO} \rightarrow Chl_{LUMO}$	0.41490	34.4%
			$Vio_{HOMO} \rightarrow Vio_{LUMO}$	0.50865	51.7%
	594 (2.09)	1.9747	$Chl_{HOMO} \rightarrow Chl_{LUMO}$	-0.41815	35.0%
			$Vio_{HOMO} \rightarrow Vio_{LUMO}$	0.46960	44.1%

Table 5.2: Overview of the details of the Chlorophyll b and Violaxanthin excitations with strong HOMO $\rightarrow$ LUMO character in the coupled systems (models 3 and 3+PCF). Characters other than HOMO $\leftrightarrow$ LUMO are omitted for clarification. The percentage of contribution of each character to an excited state is given by (coefficient)$^2 \times 200\%$.

To see if the mixing of the two HOMO $\rightarrow$ LUMO transitions holds for another state of protein protonation and overall charge, we compared the above calculations to calculations with fully protonated His residues. It was found that the coupling is slightly reduced upon His protonation, see appendix A.2.4. As the closest His residue is approximately 10 Å away from the chromophore pair, and we represent the His residues purely as point charges, the effect probably results from long-range dipole effects on the chromophores.

The fact that the CAM-B3LYP excited states of model 3 and 3+PCF consist of the HOMO $\rightarrow$ LUMO transitions of *both* chromophores is striking, even though it was found to be weaker for the monomers B and C. It implies that the two chromophores cannot be excited individually, and the two one-electron HOMO $\rightarrow$ LUMO excitations of each molecule mix to form two two-electron excitations. One requires however several external

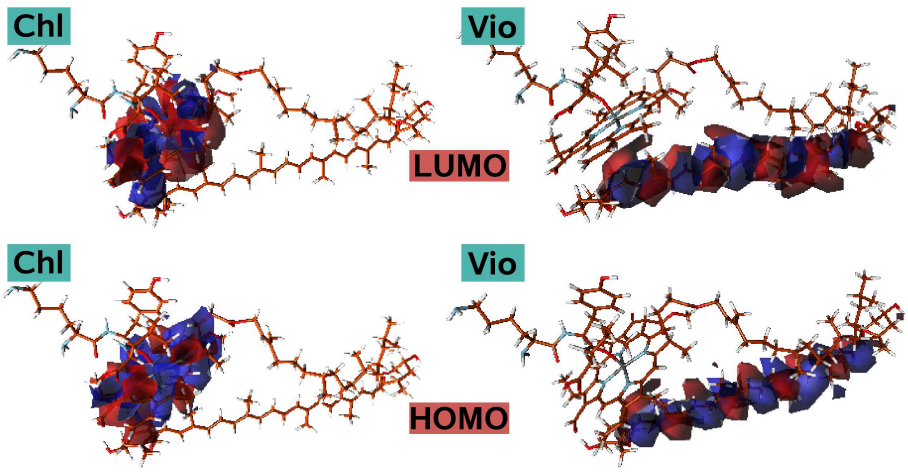

Figure 5.9: CAM-B3LYP/6-31G(d) Kohn-Sham HOMO and LUMO of Chlorophyll b and Violaxanthin for the coupled model *in vacuo* (model 3). Identification of the orbitals according to the criteria in the text. Spatial properties correspond well to the orbitals expected for isolated molecules.

constraints resulting from the protein environment (geometric restraints and Coulomb interaction), which only in combination yield the observed properties within our model calculations.

5.3 Summary and Discussion (LHCII)

It was shown that almost any change to the environment or geometry of Violaxanthin and Chlorophyll b will induce a red shift in the bands with strong HOMO $\rightarrow$ LUMO character. This effect was found in the TD-B3LYP/6-31G(d) and TD-CAM-B3LYP/6-31G(d) UV/vis spectra of both molecules, as well with CIS/6-31G(d). The red shift was observed for both Vio and Chl in all three available LHCII monomers and was more pronounced for Violaxanthin than for Chlorophyll b.

For the models 1, we have seen that B3LYP underestimates $S^{\mathrm{Vio}}_{0\rightarrow 2}$ in terms of excitation energy. In all more advanced models, it can be seen that in B3LYP spectra, $S^{\mathrm{Vio}}_{0\rightarrow 2}$ is computed to be energetically lower than $S^{\mathrm{Chl}}_{0\rightarrow 1}$. This can be considered rather unphysical, as it would hinder known mechanisms of light harvesting. Yet, since the CAM-B3LYP results for the solvated Vio and Chl was satisfying with respect to the experimental data, we can expect the observed trends to be realistic. Indeed, CAM-B3LYP predicts $S^{\mathrm{Vio}}_{0\rightarrow 2}$ to

be energetically higher than $S^{Chl}_{0\to1}$, regardless of conformation, environment or monomer. This is interesting, as one could expect $S^{Vio}_{0\to2}$ to slip below $S^{Chl}_{0\to1}$ at some point, as it is much more affected by the investigated model improvements. This, however, was not found for any case.

The actual strength of the red shift depends on the change introduced to the system: The most significant change (larger than 0.3 eV for Vio; about 0.1 eV for Chl) is seen when forcing the chromophores to take on their crystal geometry. Calculating both chromophores as a combined pair or introducing a point charge field results in red shifts smaller than 0.1 eV. It can be found, however, that the combined pair displays mixed electronic states: Both $S^{Vio}_{0\to2}$ and $S^{Chl}_{0\to1}$ bands have strong HOMO $\to$ LUMO character arising from both chromophores.

The amount of mixed HOMO $\to$ LUMO character depends on the computational method and the environment. For B3LYP, the coupling is small ($\leq$5%), regardless of an *in vacuo* model or PCF environment. For CAM-B3LYP however, the mixing within the vacuum models is already strong (16.3% Chl HOMO $\to$ LUMO in $S^{Vio}_{0\to2}$, 21.7% Vio HOMO $\to$ LUMO in $S^{Chl}_{0\to1}$ for monomer A). In the PCF environment, the mixing is even stronger, as the two bands cannot be distinguished anymore, at least in one of the three LHCII monomers. In that case, the oscillator strength is evenly distributed between the two bands. This way, the Chl HOMO $\to$ LUMO transition is strongly enhanced, because the Vio HOMO $\to$ LUMO transition practically "donates" oscillator strength to Chl. This enhancement is present in all monomers, even in the cases where the mixing is not as strong as in monomer A.

The direct mixing of the two chromophores cannot easily be observed in experimental spectra of the isolated molecules. Indeed, it is tempting to construct an alternative mechanism for light harvesting from the results shown (direct excitonic coupling). Yet, this would not rule out the relaxation to S^{Vio}_1, and it may well be that both pathways play a role in the processes of light harvesting.

Further test calculations also hinted at that the mixing is systematically decreased when fully protonating the His residues (see appendix A.2.1). This is intriguing, as most His residues are facing the luminal side - and light harvesting is supposed to stop upon lumen

acidification.

Regarding the amount of HOMO $\rightarrow$ LUMO mixing, the obvious differences between B3LYP and CAM-B3LYP give rise to the assumption that the corrected $1/r$ behavior of CAM-B3LYP is important for the effect. We have shown that CAM-B3LYP is better suited to reproduce the experimental data for Violaxanthin in acetone. Additionally, since both the intermolecular interaction and the PCF are long-range effects, the $1/r$ behavior may be expected to affect our results to a certain degree. Table 5.2 shows that the strong mixing observed with CAM-B3LYP already without the PCF can only be seen weakly with B3LYP, if at all. It is thus likely that the asymptotic corrections within CAM-B3LYP are crucial for systems as large and complex as the ones investigated here.

Test calculations (see appendix A.2.3) have shown that the phytyl chain of Chl is modulating the maximum of $S_{0\rightarrow 2}^{\mathrm{Vio}}$ as well. Future calculations might therefore be aimed at investigating the role of certain parts of the chromophores. Our calculations hint at the red shift being a systematic change induced by several factors. The phytyl chain might be one of them.
In the PDB dataset, there are also other chromophores close to Vio (a chlorophyll a and another chlorophyll b). A lipid is embedded directly next to the chromophore pair of our study, too. All these molecules might be incorporated in future investigations of similar fashion, as they are not covalently bound to the protein structure. It might be that, *e.g.*, the lipid is not only a structural element, but also required in modulating the Vio spectrum *via* intermolecular interaction.

All investigated chromophores show a distinct fine structure of the UV/vis bands in experiment. To see if our predictions hold, vibronic broadening should be applied in a future study. The vertical excitations alone are just an approximation to the real UV/vis behavior. Additionally, a MD simulation for structural sampling should be performed in the future to see if the mixing holds for other LHCII conformations, too. The difficulties arising from TD-DFT for the description of S_1^{Vio} may be overcome using more sophisticated, but computationally very demanding methods like DFT/MRCI, as has been done

in a recent study [149].

Finally, it should be mentioned that the presented calculations did not yield any insights on the process of non-photochemical quenching. But if the spectral behavior of Vio presented here proves to be correct, this might also have consequences for the research of NPQ.

Chapter 6

Conclusions

6.1 Summary of the Thesis

We were able to successfully implement geometries and methodologies for all of the biological systems in this thesis. From there, we were able to provide answers for some questions regarding each system:

- We were able to show that the **RuBisCO** carbon isotope effect can originate from the change in ZPE between reactant and transition state in the carboxylation step. The effect should then also occur regardless of the reaction center geometry and seems to be largely independent of QM method, basis set and, supposedly, environment.

- For the **BLUF** domains, we found another candidate amino acid (Ser41) for spectral influence or possible functionality. This was recently experimentally supported [127]. Our own further studies hint at the Ser41 spectral influences being of dynamical origin, so minimum geometries (like pure QM studies on optimized geometries) should not be able to grasp the full picture.

- In the **LHCII** system, we found an effect which might be a simple way to achieve light harvesting. When using the CAM-B3LYP method, the HOMO $\rightarrow$ LUMO transitions of both chromophores violaxanthin and chlorophyll b are strongly mixed, which is supported by the protein if mimicked by a point charge field. Violaxanthin donates oscillator strength to the energetically lowest, optically allowed chlorophyll b excitation, enhanced by geometric distortion and the Colomb effects of the protein environment.

All the given systems share that previously, their protein environment was either mostly neglected or approximated using PCM. Our results show that this approach may lead to nonobservance of important aspects.

We did not study the RuBisCO in a point charge field as it was the case for the other systems. However, we have shown that small changes in model composition or reaction mechanism already have drastic effects on the kinetic barriers. A direct comparison of any proposed mechanism, as we did here, may yield more insight than concentrating on a single mechanism alone. This way, we found that the isotopic effect for the CO_2 carbon atom can be produced by the carboxylation step regardless of the mechanism.
This isotopic effect is probably based simply on the C(CO_2)-C(RuBP) reaction coordinate, an aspect we would have not been able to observe if we would have employed only a single reaction mechanism.

The BLUF system was up to the date of this thesis already extensively studied. However, most models did only include amino acids which were already beforehand experimentally tested for their role. We used a "brute force approach", including all of the environment in close vicinity to the chromophore. This way, we were able to show that Ser41 may play a role at least for the UV/vis spectrum of BLUF. Taking possible interaction partners into account, it may well serve as a partial anchor for the possibly central Trp104 residue. A recent study has shown that Slr1694 BLUF mutants lacking the serine residue exhibit a simpler excitation behavior as well as a red shifted dark and light state spectrum [127].
Our ongoing computational studies revealed no dependence of the BLUF UV/vis spectrum on the position of Trp104. Yet, we find that the Ser41 hydrogen bonding is affected strongly by the Trp104 position. We suspect methodological difficulties to be the reason for this inconsistency, as the effect of Ser41 on the flavin spectrum is likely to be of dynamic, thermal character. The necessary re-optimizations of our MD snapshot geometries may result into an artificial equalization of the thermal ensembles, as they partially correspond to a structure at 0 Kelvin.

The LHCII system was studied in terms of isolated and coupled chlorophyll b and violaxanthin. Additionally, we applied a point charge field and compared the results to vacuum models. We saw that the environment drastically influences the chromophore coupling. This effect was most obvious when using CAM-B3LYP. If our results prove to be consistent with other future studies, we might have shown that not only the environment itself is important, but also the choice of method to obtain the correct interaction between environment and model system.

6.2 Future Perspectives

The influences of the protein environment on each of the systems need to be further explored. We have shown that slight changes in the model, like including more of the environment or a point charge field, can already induce severe qualitative changes. These changes can affect heights of reaction barriers, but also spectral features may shift in wavelength and/or intensity.

More specifically, for the RuBisCO system, we expect an extended model to result directly into a clearer understanding of the reaction kinetics. As yet, many features remain unresolved, most prominently the problem of RuBisCO's extremely low turnover rate [72]. Finding the rate limiting step still needs a lot of modeling work in the future, as none of the reaction pathways proposed so far is capable of explaining both the slow reaction as well as the carboxylation effect. Also, cooperative effects between the two reaction centers of an 2L complex may be explored in the future.

The BLUF system needs to be explored in a more global manner, as the conformational change of the whole domain in not yet explained. There have been reports for a possible signal transduction mechanism [108], however, the described protein lacks both the Trp and the Ser residues described here. Thus, we cannot safely transfer the findings for this BLUF domain to a general mechanism of all BLUF domains without further investigations. This, of course, also applies for our results. As a consequence, analogous to the RuBisCO case, several types of BLUF domains have to be mechanistically compared

before safe statements can be made about either of them.

Finally, the LHCII complex is far from being completely understood. Current computational resources forbid the explicit treatment of more than a few chromophores within a single TD-DFT calculation. Thus, the investigation of large, coupled chromophore complexes like LHCII strongly depend on the available computers. LHCII is methodologically less demanding than other systems, since none of the chromophores is covalently bonded to the protein. Simple point charge field schemes, as we applied them in this thesis, may suffice for an accurate treatment of the protein environment. However, as long as computational resources forbid larges systems, we have to restrict the studies on one or two coupled chromophores.

We have seen that in all discussed systems, including the environment may or may not have strong influence on the qualitative results. Especially the BLUF case has shown that it is easy to overlook an important aspect if one does not include almost all possible influences from the start. There is, however, no way to know the environment's influence beforehand.
The usual approach of slowly increasing model complexity is definitely necessary for the final understanding of the system. For weakly bonded chromophores, like in the LHCII case, this approach might actually be more successful. Yet, for initial exploration of a strongly bonded chromophore, a "brute force" approach may be required. Including most of the immediate environment might be better as initial model, as it generates an overview on all participating structures. From there, it is possible to develop smaller systems for the standard "bottom up" approach. This is a fundamental difference to usual quantum chemical procedures.
This thesis is concluded with the statement that the effects arising from a complex protein environment should not be underestimated.

Bibliography

[1] Kitano, H. Systems biology: A brief overview. *Science* **295**, 1662–1664 (2002).

[2] Ghaemmaghami, S., Huh, W., Bower, K., Howson, R. W., Belle, A., Dephoure, N., O'Shea, E. K. & Weissman, J. S. Global analysis of protein expression in yeast. *Nature* **425**, 737–741 (2003).

[3] Boccaletti, S., Latora, V., Moreno, Y., Chavez, M. & Hwang, D. U. Complex networks: Structure and dynamics. *Phys. Rep.* **424**, 175–308 (2006).

[4] Alhambra, C., Corchado, J. C., Sanchez, M. L., Gao, J. L. & Truhlar, D. G. Quantum dynamics of hydride transfer in enzyme catalysis. *J. Am. Chem. Soc.* **122**, 8197–8203 (2000).

[5] Wang, W., Donini, O., Reyes, C. M. & Kollman, P. A. Biomolecular simulations: Recent developments in force fields, simulations of enzyme catalysis, protein-ligand, protein-protein, and protein-nucleic acid noncovalent interactions. *Annu. Rev. Bioph. Biom.* **30**, 211–243 (2001).

[6] Field, M. J. Simulating enzyme reactions: Challenges and perspectives. *J. Comp. Chem.* **23**, 48–58 (2002).

[7] Hartree, D. R. The wave mechanics of an atom with a non-Coulomb central field Part I Theory and methods. *P. Camb. Philos. Soc.* **24**, 89–110 (1928).

[8] Fock, V. A. Näherungsmethode zur Lösung des quantenmechanischen Mehrkörperproblems. *Z. Phys.* **61**, 126–147 (1930).

[9] Hohenberg, P. C. & Kohn, W. Inhomogenous electron gas. *Phys. Rev. B* **136**, B864–B871 (1964).

[10] Pullman, B. & Pullman, A. Electron-donor and electron acceptor properties of biologically important purines, pyrimidines, pteridines, flavins, and aromatic amino acids. *P. Natl. Acad. Sci. USA* **44**, 1197–1202 (1958).

[11] Dewar, M. J. S. & Thiel, W. Ground-states of molecules .38. Mndo method - Approximations and parameters. *J. Am. Chem. Soc.* **99**, 4899–4907 (1977).

[12] Brooks, B. R. *et al.* CHARMM - A program for macromolecular energy, minimization, and dynamics calculations. *J. Comp. Chem.* **4**, 187–217 (1983).

[13] Case, D. *et al.* Amber 11. *Univ. of California, SF* (2010).

[14] Scott, W. R. P. *et al.* The GROMOS biomolecular simulation program package. *J. Phys. Chem. A* **103**, 3596–3607 (1996).

[15] Senn, H. S. & Thiel, W. QM/MM methods for biomolecular systems. *Angew. Chem. Int. Edit.* **48**, 1198–1229 (2009).

[16] Frank, H. A., Cua, A., Chynwat, V., Young, A., Gosztola, D. & Wasielewski, M. R. Photophysics of the carotenoids associated with the xanthophyll cycle in photosynthesis. *Photosynth. Res.* **41**, 389–395 (1994).

[17] Stitt, M. Rising CO2 levels and their potential significance for carbon flow in photosynthetic cells. *Plant Cell Environ.* **14**, 741–762 (1991).

[18] Roeske, C. A. & O'Leary, M. H. Carbon isotope effects on the enzyme-catalyzed carboxylation of ribulose bisphosphate. *Biochemistry-US* **23**, 6275–6284 (1984).

[19] Gomelsky, M. & Klug, G. BLUF: A novel FAD-binding domain involved in sensory transduction in microorganisms. *Trends Biochem. Sci.* **27**, 497–500 (2002).

[20] Jensen, F. Introduction to computational chemistry, 2nd edition, english. *Wiley & Sons* (2007).

[21] Atkins, P. & Friedman, R. Molecular quantum dynamics, 4th edition, english. *Oxford University Press, New York* (2005).

[22] Dewar, M. J. S., Zoebisch, E. G., Healy, E. F. & Stewart, J. J. P. The development and use of quantum-mechanical molecular-models. 76. AM1 - A new general-purpose quantum-mechanical molecular-model. *J. Am. Chem. Soc.* **107**, 3 (1985).

[23] Stewart, J. J. P. Optimization of parameters for semiempirical methods. 1. Method. *J. Comp. Chem.* **10**, 209–220 (1989).

[24] Sherrill, C. D. & Schaefer, H. F. The configuration interaction method: Advances in highly correlated approaches. *Adv. Quant. Chem.* **34**, 143–269 (1999).

[25] Bartlett, R. J. Coupled-cluster approach to molecular-structure and spectra - A step toward predictive quantum-chemistry. *J. Phys. Chem.* **93**, 1697–1708 (1989).

[26] Kohn, W. & Sham, L. J. Quantum density oscillations in an inhomogeneous electron gas. *Phys. Rev.* **137**, 1697 (1965).

[27] Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* **38**, 3098–3100 (1988).

[28] Perdew, J. P. Density-functional approximation for the correlation energy of the inhomogenous electron gas. *Phys. Rev. B* **33**, 8822–8824 (1986).

[29] Becke, A. D. Density-functional thermochemistry. 3. The role of exact exchange. *J. Chem. Phys.* **98**, 5648–5652 (1993).

[30] Stephens, P. J., Devlin, F. J., Chabalowski, C. F. & Frisch, M. J. Ab-initio calculation of vibrational absorption and circular-dichroism spectra using density-functional force-fields. *J. Phys. Chem.* **98**, 11623–11627 (1994).

[31] Lee, C., Yang, W. & Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron-density. *Phys. Rev. B* **37**, 785–789 (1988).

[32] Yanai, T., Tew, D. & Handy, N. A new hybrid exchange-correlation functional using the Coulomb-attenuating method (CAM-B3LYP). *Chem. Phys. Lett.* **393**, 51–57 (2004).

[33] Foresman, J. B., Head-Gordon, M., Pople, J. A. & Frisch, M. J. Toward a systematic molecular orbital theory for excited states. *J. Phys. Chem.* **96**, 135–149 (1992).

[34] Jamorski, C., Casida, M. E. & Salahub, D. R. Dynamic polarizabilities and excitation spectra from a molecular implementation of time-dependent density-functional response theory: N_2 as a case study. *J. Chem. Phys.* **104**, 5134–5147 (1996).

[35] Casida, M. E. Recent Advances in Density Functional Methods. *D. P. Chong ed., World Scientific*, 155–192 (1995).

[36] Scott, A. P. & Radom, L. Harmonic vibrational frequencies: An evaluation of Hartree-Fock, Møller-Plesset, quadratic configuration interaction, density functional theory, and semiempirical scale factors. *J. Phys. Chem.* **100**, 16502–16513 (1996).

[37] Schultz, N. E., Zhao, Y. & Truhlar, D. G. Databases for transition element bonding: Metal-metal bond energies and bond lengths and their use to test hybrid, hybrid meta, and meta density functionals and generalized gradient approximations. *J. Phys. Chem. A* **109**, 4388–4403 (2005).

[38] Halgren, T. A. & Lipscomb, W. N. The synchronous-transit method for determining reaction pathways and locating molecular transition states. *Chem. Phys. Lett.* **49**, 225–232 (1977).

[39] Miertus, S., Scrocco, E. & Tomasi, J. Electrostatic interaction of a solute with a continuum - A direct utilization of ab initio molecular potentials for the prevision of solvent effects. *Chem. Phys.* **55**, 117–129 (1981).

[40] Frisch, M. J. *et al.* Gaussian 09, Revision A.02. *Gaussian, Inc., Wallingford CT* (2004).

[41] Mennucci, B., Cances, E. & Tomasi, J. Evaluation of solvent effects in isotropic and anisotropic dielectrics and in ionic solutions with a unified integral equation method: Theoretical bases, computational implementation, and numerical applications. *J. Phys. Chem. B* **101**, 10506–10517 (1997).

[42] Singh, U. C. & Kollman, P. A. An approach to computing electrostatic charges for molecules. *J. Comp. Chem.* **5**, 129–145 (1984).

[43] Besler, B. H., Merz Jr., K. M. & Kollman, P. A. Atomic charges derived from semiempirical methods. *J. Comp. Chem.* **11**, 431–439 (1990).

[44] Strain, M. C., Scuseria, G. E. & Frisch, M. J. Achieving linear scaling for the electronic quantum Coulomb problem. *Science* **271**, 51–53 (1996).

[45] Bakowies, D. & Thiel, W. Hybrid models for combined quantum mechanical and molecular mechanical approaches. *J. Phys. Chem.* **100**, 10580–10594 (1996).

[46] Vreven, T., Komaromi, I., Byun, K. S., Dapprich, S., Montgomery, J. A., Morokuma, K. & Frisch, M. J. Combining quantum mechanics methods with molecular mechanics methods in ONIOM. *J. Chem. Theory Comput.* **2**, 815–826 (2006).

[47] Patel, S., MacKerell, A. D. & Brooks, C. CHARMM fluctuating charge force field for proteins: II - Protein/solvent properties from molecular dynamics simulations using a nonadditive electrostatic model. *J. Comp. Chem.* **25**, 1504–1514 (2004).

[48] Ren, P. Y. & Ponder, J. W. Polarizable atomic multipole water model for molecular mechanics simulation. *J. Phys. Chem. B* **107**, 5933–5947 (2003).

[49] Bekker, H. *et al.* GROMACS: A parallel computer for molecular dynamics simulations. *Physics Computing 92, Singapore, de Groot, R. A., Nadrchal, J. eds.* World Scientific (1993).

[50] MacKerell, A. D., Banavali, N. & Foloppe, N. Development and current status of the CHARMM force field for nucleic acids. *Biopolymers* **56**, 257–265 (2000).

[51] Wang, J. M., Cieplak, P. & Kollman, P. A. How well does a restrained electrostatic potential (resp) model perform in calculating conformational energies of organic and biological molecules? *J. Comp. Chem.* **21**, 1049–1074 (2000).

[52] ambermd.org/eqn.txt. Amber official website .

[53] Sorin, E. J. & Pande, V. S. Exploring the helix-coil transition via all-atom equilibrium ensemble simulations. *Biophys. J.* **88**, 2472–2493 (2005).

[54] Swope, W. C., Andersen, H. C., Berens, P. H. & Wilson, K. R. A computer-simulation method for the calculation of equilibrium-constants for the formation of

physical clusters of molecules - application to small water clusters. *J. Chem. Phys.* **76**, 637–649 (1982).

[55] Hockney, R. W., Goel, S. P. & Eastwood, J. Quiet high-resolution computer models of a plasma. *J. Comp. Phys.* **14**, 148–158 (1974).

[56] Verlet, L. Computer experiments on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. *Phys. Rev.* **159**, 98–103 (1967).

[57] Berendsen, H. J. C. Transport properties computed by linear response through weak coupling to a bath. *In: Computer Simulations in Material Science. Meyer; Pontikis, V. eds., Kluwer* 139–155 (1991).

[58] Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126**, 014101 (2007).

[59] Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An n-log(n) method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089–10092 (1993).

[60] Dapprich, S., Komaromi, I., Byun, K. S., Morokuma, K. & Frisch, M. J. A new ONIOM implementation in Gaussian 98. 1. The calculation of energies, gradients and vibrational frequencies and electric field derivatives. *J. Mol. Struc.-THEOCHEM* **462**, 1–21 (1999).

[61] Ryde, U. & Olsson, M. H. M. Structure, strain, and reorganization energy of blue copper models in the protein. *Intern. J. Quant. Chem.* **81**, 335–347 (2001).

[62] TURBOMOLE V6.1 2009, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989-2007, TURBOMOLE GmbH, since 2007; available from `http://www.turbomole.com`.

[63] Andersson, I. & Backlund, A. Structure and function of RuBisCO. *Plant Phys. Biochem.* **46**, 275–291 (2007).

[64] Tcherkez, G. B. G., Farquhar, G. D. & Andrews, T. J. Despite slow catalysis and confused substrate specificity, all ribulose bisphosphate carboxylases may be nearly perfectly optimized. *P. Natl. Acad. Sci. USA* **103**, 7246–7251 (2006).

[65] Cleland, W. W., Andrews, T. J., Gutteridge, S., Hartman, F. C. & Lorimer, G. H. Mechanism of RuBisCO: The carbamate as general base. *Chem. Rev.* **98**, 549–561 (1998).

[66] Jensen, R. G. & Bahr, J. T. Ribulose 1,5-bisphosphate carboxylase-oxygenase. *Ann. Rev. Plant Physo.* **28**, 379–400 (1977).

[67] Portis, A. R. & Parry, M. A. J. Discoveries in RuBisCO (ribulose 1,5-bisphosphate carboxylase/oxygenase): A historical perspective. *Photosynth. Res.* **94**, 121–143 (2007).

[68] Andersson, I. Catalysis and regulation in RuBisCO. *J. Exp. Bot.* **59**, 1555–1568 (2008).

[69] Heldt, H. W. Pflanzenbiochemie, 2nd edition, German. *Spektrum Akademischer Verlag Heidelberg* (1999).

[70] Taylor, T. C. & Andersson, I. Structure of a product complex of spinach ribulose-1,5-bisphosphate carboxylase/oxygenase. *Biochemistry-US* **36**, 4041–4046 (1997).

[71] Götze, J. & Saalfrank, P. Quantum chemical modelling of the RuBisCO carbon isotope effect. *Submitted* (2010).

[72] Farquhar, G. D. Models describing the kinetics of ribulose biphosphate carboxylase-oxygenase. *Arch. Biochem. Biophys.* **193**, 456–468 (1979).

[73] Pearce, F. G. Catalytic by-product formation and ligand binding by ribulose bisphosphate carboxylases from different phylogenies. *Biochem. J.* **399**, 525–534 (2006).

[74] Witzel, F., Götze, J. & Ebenhoeh, O. Slow deactivation of RuBisCO elucidated by mathematical models. *FEBS J.* **277**, 931–950 (2010).

[75] Taylor, T. C., Backlund, A., Bjorhall, K., Spreitzer, R. J. & Andersson, I. First crystal structure of RuBisCO from a green alga, *Chlamydomonas reinhardtii*. *J. Biol. Chem.* **276**, 48159–48164 (2001).

[76] Kannappan, B. & Gready, J. E. Redefinition of RuBisCO carboxylase reaction reveals origin of water for hydration and new roles for active-site residues. *J. Am. Chem. Soc.* **130**, 15063–15080 (2008).

[77] Duff, A. P., Andrews, T. J. & Curmi, P. M. G. The transition between the open and closed states of RuBisCO is triggered by the inter-phosphate distance of the bound bisphosphate. *J. Mol. Biol.* **298**, 903–916 (2000).

[78] Lee, G. J., McDonald, K. A. & McFadden, B. A. Leucine-332 influences the CO_2/O_2 specificity factor of ribulose-1,5-bisphosphate carboxylase oxygenase from *Anacystis-nidulans*. *Prot. Sci.* **2**, 1147–1154 (1993).

[79] Pearce, F. G. & Andrews, T. J. The relationship between side reactions and slow inhibition of ribulose-bisphosphate carboxylase revealed by a loop 6 mutant of the tobacco enzyme. *J. Biol. Chem.* **278**, 32526–32536 (2003).

[80] Mauser, H., King, W. A., Gready, J. E. & Andrews, T. J. CO_2 fixation by RuBisCO: Computational dissection of the key steps of carboxylation, hydration, and C-C bond cleavage. *J. Am. Chem. Soc.* **123**, 10821–10829 (2001).

[81] Mizohata, E., Matsumura, H., Okano, Y., Kumei, M., Takuma, H., Onodera, J., Kato, K., Shibata, N., Inoue, T., Yokota, A. & Kai, Y. Crystal structure of activated ribulose-1,5-bisphosphate carboxylase/oxygenase from green alga *Chlamydomonas reinhardtii* complexed with 2-carboxyarabinitol-1,5-bisphosphate. *J. Mol. Biol.* **316**, 679–691 (2002).

[82] Sousa, S. F., Fernandes, P. A. & Ramos, M. J. The carboxylate shift in zinc enzymes: A computational study. *J. Am. Chem. Soc.* **129**, 1378–1385 (2007).

[83] Tcherkez, G. & Farquhar, G. D. Carbon isotope effect predictions for enzymes involved in the primary carbon metabolism of plant leaves. *Func. Plant Biol.* **32**, 227–291 (2005).

[84] Gomelsky, M. & Kaplan, S. AppA, a redox regulator of photosystem formation in rhodobacter sphaeroides 2.4.1, is a flavoprotein - Identification of a novel FAD binding domain. *J. Biol. Chem.* **273**, 35319–25325 (1998).

[85] Braatsch, S., Gomelsky, M., Kuphal, S. & Klug, G. A single flavoprotein, AppA, integrates both redox and light signals in *Rhodobacter sphaeroides*. *Mol. Microbiol.* **45**, 827–836 (2002).

[86] Hegemann, P. Algal sensory photoreceptors. *Ann. Rev. Plant Biol.* **59**, 167–189 (2008).

[87] Laan, W., Bednarz, T., Heberle, J. & Hellingwerf, K. J. Chromophore composition of a heterologously expressed BLUF-domain. *Photochem. Photobiol. Sci.* **3**, 1011–1016 (2004).

[88] Zirak, P., Penzkofer, A., Schiereis, T., Hegemann, P. & Jung, I., A. Schlichting. Absorption and fluorescence spectroscopic characterization of BLUF domain of AppA from *Rhodobacter sphaeroides*. *Chem. Phys.* **315**, 142–154 (2005).

[89] Neiss, C., Saalfrank, P., Parac, M. & Grimme, S. Quantum chemical calculation of excited states of flavin-related molecules. *J. Phys Chem. A* **107**, 140–147 (2003).

[90] Neiss, C. & Saalfrank, P. Ab initio quantum chemical investigation of the first steps of the photocycle of phototropin: A model study. *Photochem. Photobiol.* **77**, 101–109 (2003).

[91] Zenichowski, K., Gothe, M. & Saalfrank, P. Exciting flavins: Absorption spectra and spin-orbit coupling in light-oxygen-voltage (LOV) domains. *Photochem. Photobiol. A* **190**, 290–300 (2007).

[92] Masuda, S., Hasegawa, K., Ishii, A. & Ono, T. Light-induced structural changes in a putative blue-light receptor with a novel fad binding fold sensor of blue-light using FAD (BLUF); Slr1694 of *Synechocystis sp. PCC6803*. *Biochemistry-US* **43**, 5304–5313 (2004).

[93] Domratcheva, T., Grigorenko, B. L., Schlichting, I. & Nemukhin, A. V. Molecular models predict light-induced glutamine tautomerization in BLUF photoreceptors. *Biophys. J.* **94**, 3872–3879 (2008).

[94] Sadeghian, K., Bocola, M. & Schütz, M. A conclusive mechanism of the photoinduced reaction cascade in blue light using flavin photoreceptors. *J. Am. Chem. Soc.* **130**, 12501–12513 (2008).

[95] Salzmann, S., Tatchen, J., & Marian, C. M. The photophysics of flavins: What makes the difference between gas phase and aqueous solution? *J. Photochem. Photobiol. A* **198**, 221–231 (2008).

[96] Götze, J. & Saalfrank, P. Serine in BLUF domains displays spectral importance in computational models. *J. Photochem. Photobiol B* **94**, 87–95 (2009).

[97] Salzmann, S., Silva-Junior, M. R., Thiel, W. & Marian, C. M. Influence of the LOV domain on low-lying excited states of flavin: A combined quantum-mechanics/molecular-mechanics investigation. *J. Phys. Chem. B* **113**, 15610–15618 (2009).

[98] Salzmann, S., Martinez-Junza, V., Zorn, B., Braslavsky, S. E., Mansurova, M., Marian, C. M. & Gärtner, W. Photophysical Properties of structurally and electronically modified flavin derivatives determined by spectroscopy and theoretical calculations. *J. Phys. Chem. A* **113**, 9365–9375 (2009).

[99] Gauden, M., Grinstead, J. S., Laan, W., van Stokkum, H. M., Avila-Perez, M., Toh, K. C., Boehlens, R., Kaptein, R., von Grondelle, R., Hellingwerf, K. J. & Kennis, J. T. M. On the role of aromatic side chains in the photoactivation of BLUF domains. *Biochemistry-US* **46**, 7405–7415 (2007).

[100] Grinstead, J. S., Hsu, S. T. D., Laan, W., Bonvin, A. M. J. J., Hellingwerf, K. J., Boehlens, R. & Kaptein, R. The solution structure of the AppA BLUF domain: Insight into the mechanism of light-induced signaling. *ChemBioChem* **7**, 187–193 (2006).

[101] Anderson, S., Dragnea, V., Masuda, S., Ybe, J., Moffat, K. & Bauer, C. Structure of a novel photoreceptor, the BLUF domain of AppA from *Rhodobacter sphaeroides*. *Biochemistry-US* **44**, 7998–8005 (2005).

[102] Jung, A., Domratcheva, T., Tarutina, M., Wu, Q., Ko, W. H., Shoeman, R. L., Gomelsky, M., Gardner, K. H. & Schlichting, I. Structure of a bacterial BLUF photoreceptor: Insights into blue light-mediated signal transduction. *J. Mol. Biol.* **102**, 12350–12355 (2005).

[103] Kita, A., Okajima, K., Morimoto, Y., Ikeuchi, M. & Miki, K. Structure of a cyanobacterial BLUF protein, Tll0078, containing a novel FAD-binding blue light sensor domain. *J. Mol. Biol.* **349**, 1–9 (2005).

[104] Jung, A., Reinstein, J., Domratcheva, T., Shoeman, R. L. & Schlichting, I. Crystal structures of the AppA BLUF domain photoreceptor provide insights into blue light-mediated signal transduction. *J. Mol. Biol.* **362**, 716–732 (2006).

[105] Yuan, H., Anderson, S., Masuda, S., Dragnea, V., Moffat, K. & Bauer, C. Crystal structures of the *Synechocystis* photoreceptor Slr1694 reveal distinct structural states related to signaling. *Biochemistry-US* **45**, 12687–12694 (2006).

[106] Masuda, S., Hasegawa, K. & Ono, T. A. Tryptophan at position 104 is involved in transforming light signal into changes of beta-sheet structure for the signaling state in the BLUF domain of AppA. *Plant Cell Phys.* **46**, 1894–1901 (2005).

[107] Masuda, S., Hasegawa, K. & Ono, T. A. Light-induced structural changes of apoprotein and chromophore in the sensor of blue light using FAD (BLUF) domain of AppA for a signaling state. *Biochemistry-US* **44**, 1215–1224 (2005).

[108] Bahrends, T. R. M., Hartmann, E., Griese, J. J., Beitlich, T., Kirienko, N. V., Ryjenkov, D. A., Reinstein, J., Shoeman, R. L., Gomelsky, M. & Schlichting, I. Structure and mechanism of a bacterial light-regulated cyclic nucleotide phosphodiesterase. *Nature* **459**, 1015–1018 (2009).

[109] Wu, Q. & Gardner, K. H. Structure and insight into blue light-induced changes in the BlrP1 BLUF domain. *Biochemistry-US* **48**, 2620–2629 (2009).

[110] Toh, K. C., van Stokkum, I. H. M., Hendriks, J., Alexandre, M. T. A., Arents, J. C., Perez, M. A., van Grondelle, R., Hellingwerf, K. J. & Kennis, J. T. M. On

the signaling mechanism and the absence of photoreversibility in the AppA BLUF domain. *Biophys. J.* **95**, 312–321 (2008).

[111] Götze, J., Greco, C., Bonačić-Koutecký, V. & Saalfrank, P. BLUF hydrogen network dynamics explored by comparative molecular dynamics of AppA and BlrB. *in preparation* (2010).

[112] Laan, W., van der Horst, M. A., van Stokkum, I. H. & Hellingwerf, K. J. Initial characterization of the primary photochemistry of AppA, a blue-light-using flavin adenine dinucleotide-domain containing transcriptional antirepressor protein from rhodobacter sphaeroides: A key role for reversible intramolecular proton transfer from the flavin adenine dinucleotide chromophore to a conserved tyrosine? *Photochem. Photobiol.* **78**, 290–297 (2003).

[113] Masuda, S., Tomida, Y., Ohta, H. & Takamiya, K. I. The critical role of a hydrogen bond between Gln63 and Trp104 in the blue-light sensing BLUF domain that controls AppA activity. *J. Mol. Biol.* **368**, 1223–1230 (2007).

[114] Stelling, A. L., Ronayne, K. L., Nappa, J., Tonge, P. J. & Meech, S. R. Ultrafast structural dynamics in BLUF domains: Transient infrared spectroscopy of AppA and its mutants. *J. Am. Chem. Soc.* **129**, 15556–15564 (2007).

[115] Schaftenaar, G. & Noordik, J. Molden: A pre- and post-processing program for molecular and electronic structures. *J. Comput. Aid. Mol. Des.* **14**, 123–134 (2000).

[116] Klaumünzer, B., Kroener, D. & Saalfrank, P. (TD-)DFT calculation of vibrational and vibronic spectra of riboflavin in solution. *J. Phys. Chem. B* **114**, 10826–10834 (2010).

[117] Salzmann, S. & Marian, C. M. The photophysics of alloxazine: a quantum chemical investigation in vacuum and solution. *Photochem. Photobiol. Sci.* **8**, 1655–1666 (2009).

[118] Foloppe, N. & MacKerell, A. D. All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data. *J. Comput. Chem.* **21**, 86–104 (2000).

[119] Jørgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983).

[120] Dittrich, M., Freddolino, P. L. & Schulten, K. When light falls in LOV: A quantum mechanical/molecular mechanical study of photoexcitation in Phot-LOV1 of *Chlamydomonas reinhardtii*. *J. Phys. Chem. B* **109**, 13006–13013 (2005).

[121] Casida, M. E. & Salahub, D. R. Asymptotic correction approach to improving approximate exchange-correlation potentials: Time-dependent density-functional theory calculations of molecular excitation spectra. *J. Chem. Phys.* **113**, 8918–8935 (2000).

[122] Kraft, B. J., Masuda, S., Kikuchi, J., Dragnea, V., Tollin, G., Zaleski, J. M. & Bauer, C. E. Spectroscopic and mutational analysis of the blue-light photoreceptor AppA: A novel photocycle involving flavin stacking with an aromatic amino acid. *Biochemistry* **42**, 6726–6734 (2003).

[123] Hess, B., Kutzner, C., van der Spoel, D. & Lindahl, E. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theor. Comp.* **4**, 435–447 (2008).

[124] Schneider, C. & Suhnel, J. A molecular dynamics simulation of the flavin mononucleotide-RNA aptamer complex. *Biopolymers* **50**, 287 (1999).

[125] Darden, T., Perera, L., Li, L. & Pedersen, L. New tricks for modelers from the crystallography toolkit: The particle mesh Ewald algorithm and its use in nucleic acid simulations. *Structure* **7**, R55–R60 (1999).

[126] Obanayama, K., Kobayashi, H., Fukushima, K. & Sakurai, M. Structures of the chromophore binding sites in BLUF domains as studied by molecular dynamics and quatum chemical calculations. *Photochem. Photobiol.* **84**, 1003–1010 (2008).

[127] Bonetti, C., Stierl, M., Mathes, T., van Stokkum, I. H. M., Mullen, K. M., Cohen-Stuart, T. A., van Grondelle, R., Hegemann, P. & Kennis, J. T. M. The role of

key amino acids in the photoactivation pathway of the synechocystis Slr1694 BLUF domain. *Biochemistry-US* **48**, 11458–11469 (2009).

[128] Takahashi, R., Okajima, K., Suzuki, H., Nakamura, H., Ikeuchi, M. & Noguchi, T. FTIR study on the hydrogen bond structure of a key tyrosine residue in the flavin-binding blue light sensor TePixD from *Thermosynechococcus elongatus*. *Biochemistry-US* **46**, 6459–6467 (2007).

[129] Fukushima, Y., Okajima, K., Ikeuchi, M. & Itoh, S. Two intermediate states i and j trapped at low temperature in the photocycles of two BLUF domain proteins of cyanobacteria *Synechocystis sp. PCC6803* and *Thermosynechococcus elongatus BP-1*. *Photochem. Photobiol.* **83**, 112–121 (2007).

[130] Fukushima, Y., Murai, Y., Okajima, K., Ikeuchi, M. & Itoh, S. Photoreactions of Tyr8- and Gln50-mutated BLUF domains of the pixd protein of *Thermosynechococcus elongatus BP-1*: Photoconversion at low temperature without Tyr8. *Biochemistry-US* **47**, 660–669 (2008).

[131] Zhang, R. B., Somers, K. R. F., Kryachko, E. S., Nguyen, M. T., Zeegers-Huyskens, T. S. & Ceulemans, A. Hydrogen bonding to π-systems of indole and 1-methylindole: Is there any OH$\cdots$phenyl bond? *J. Phys. Chem.* **109**, 8028–8034 (2005).

[132] Standfuss, J., Terwisscha Van Scheltinga, A. C., Lamborghini, M. & Kuehlbrandt, W. Mechanisms of photoprotection and nonphotochemical quenching in pea light-harvesting complex at 2.5 a resolution. *Embo J.* **24**, 919–928 (2005).

[133] Dreuw, A., Fleming, G. R. & Head-Gordon, M. Chlorophyll fluorescence quenching by xanthophylls. *Phys. Chem. Chem. Phys.* **5**, 3247–3256 (2003).

[134] Dreuw, A. Influence of geometry relaxation on the energies of the S_1 and S_2 states of violaxanthin, zeaxanthin, and lutein. *J. Phys. Chem. A* **110**, 4592–4599 (2006).

[135] Starcke, J. H., Wormit, M., Schirmer, J. & Dreuw, A. How much double excitation character do the lowest excited states of linear polyenes have? *Chem. Phys.* **329**, 39–49 (2006).

[136] Dreuw, A. & Wormit, M. Simple replacement of violaxanthin by zeaxanthin in LHCII does not cause chlorophyll fluorescence quenching. *J. Inorg. Biochem.* **102**, 458–465 (2008).

[137] Wormit, M., Harbach, P. H. P., Mewes, J. M., Amarie, S., Wachtveitl, J. & Dreuw, A. Excitation energy transfer and carotenoid radical cation formation in light harvesting complexes - a theoretical perspective. *BBA-Bioenergetics* **1787**, 738–746 (2009).

[138] Ostroumov, E., Müller, M. G., Marian, C. M., Kleinschmidt, M. & Holzwarth, A. R. Electronic coherence provides a direct proof for energy-level crossing in photoexcited lutein and beta-carotene. *Phys. Rev. Lett.* **103**, 108302 (2009).

[139] Mueller, P., Xiao-Ping, L. & Niyogi, K. K. Non-photochemical quenching. A response to excess light energy. *Plant Physiol.* **125**, 1558–1566 (2001).

[140] Bugos, R. C. & Yamamoto, H. Y. Molecular cloning of violaxanthin de-epoxidase from romaine lettuce and expression in *Escherichia coli*. *P. Natl. Acad. Sci. USA* **93**, 6320–6325 (1996).

[141] Bouvier, F., d'Harlingue, A., Hugueney, P., Marin, E., Marion-Poll, A. & Camara, B. Xanthophyll biosynthesis - cloning, expression, functional reconstitution, and regulation of beta-cyclohexenyl carotenoid epoxidase from pepper (*Capsicum annuum*). *J. Biol. Chem.* **271**, 28861–28867 (1996).

[142] Polivka, T. & Sundstroem, V. Ultrafast dynamics of carotenoid excited states - from solution to natural and artificial systems. *Chem. Rev.* **104**, 2021–2071 (2004).

[143] Duffy, C. D. P., Ruban, A. V. & Barford, W. Theoretical investigation of the role of strongly coupled chlorophyll dimers in photoprotection of LHCII. *J. Phys. Chem. B* **112**, 12508–12515 (2008).

[144] Eskling, M., Arvidsson, P. O. & Akerlund, H. E. The xanthophyll cycle, its regulation and components. *Physiol. Plantarum* **100**, 806–816 (1997).

[145] Stewart, J. J. P. Optimization of parameters for semiempirical methods V: Modification of NDDO approximations and application to 70 elements. *J. Mol. Model.* **13**, 1173–1213 (2007).

[146] Gouterman, M. Spectra of porphyrins. *J. Mol. Spectrosc.* **68** (1961).

[147] Various authors: Phytoplankton Pigments in oceanography: Guidelines to modern methods. *Oceanographic Methodology series; Jeffrey, S. W., Mantoura, R. F. C., Wright, S. W., eds.* UNESCO Publishing (1997).

[148] Renstrom, B., Borch, G., Skulberg, O. M. & Liaaen-Jensen, S. Natural occurrence of enantiomeric and meso-astaxanthin. 3. Optical purity of (3S,3'S)-astaxanthin *ex haematococcus pluvialis* (green algae). *Phytochem.* **20**, 2561–2564 (1981).

[149] Ceron-Carrasco, J. P., Requena, A. & Marian, C. M. Theoretical study of the low-lying excited states of beta-carotene isomers by a multireference configuration interaction method. *Chem. Phys.* **373**, 98–103 (2010).

Appendix A

A.1 Details on RuBisCO Results

A.1.1 6-31G(d) Test Calculations for the New Model

We conducted a basis set test with the 6-31G(d) basis instead of 6-31G(d,p) for the new model [76]. All calculations *in vacuo*, the results are displayed in table A.1.

Method	HF/6-31G(d)	B3LYP/6-31G(d)
$\Delta G^{\ddagger}_{12}$	77.087	36.468
ΔG_{12}	-36.662	-9.363
k_{12}	$1.94{\cdot}10^{-1}$ s^{-1}	$2.54{\cdot}10^{6}$ s^{-1}
$\Delta G^{\ddagger}_{13}$	77.203	36.563
ΔG_{13}	-36.686	-9.357
k_{13}	$1.85{\cdot}10^{-1}$ s^{-1}	$2.44{\cdot}10^{6}$ s^{-1}
α	1.048	1.039

Table A.1: Summary of the carboxylation Gibbs free activation and reaction energies (from scaled frequencies) and for different isotopes (^{12}C and ^{13}C) within the new model, see eqs. (3.2) and (3.3). Without and with PCM ($\epsilon = 4.7113$; chloroform). Kinetic rates derived thereof are also shown, obtained using eq. (2.67). All energies are in kJ/mol. The isotope effect α is calculated using eq. (3.1).

A.1.2 ZPE values

To show the origin of the isotopic effect, the ΔZPE values for the calculated models are listed in tables A.2 and A.3. The ΔZPE values were calculated from $\mathrm{ZPE_{ts}} - \mathrm{ZPE_{reactants}}$.

Method	HF/6-31G(d,p)		
Model (atoms)	1 (102)	1b (19)	1 PCM (102)
$\Delta\mathrm{ZPE}_{12}$	-3.182	-0.998	-5.957
$\Delta\mathrm{ZPE}_{13}$	-3.109	-0.901	-5.876
ΔΔZPE	0.073 (71.3%)	0.097 (80.3%)	0.081 (77.1%)

Table A.2: ΔZPE values for the conventional model, in kJ/mol. Indexed are corresponding isotopes. Percentage indicates contribution of ΔΔZPE to $(\Delta G_{13} - \Delta G_{12})$ (see tables 3.1 and 3.2).

Method	HF/6-31G(d,p)		B3LYP/6-31G(d,p)	
Environment	Vacuum	PCM	Vacuum	PCM
$\Delta\mathrm{ZPE}_{12}$	3.103	0.601	5.461	4.710
$\Delta\mathrm{ZPE}_{13}$	3.193	0.701	5.535	4.799
ΔΔZPE	0.090 (76.2%)	0.100 (77.7%)	0.074 (85.4%)	0.089 (75.3%)

Table A.3: ΔZPE values for the new model, in kJ/mol. Indexed are corresponding isotopes. Percentage indicates contribution of ΔΔZPE to $(\Delta G_{13} - \Delta G_{12})$ (see tables 3.1 and 3.2).

Note that in the conventional model, the ZPE of the TS is lowered less in the case of ^{13}C. In contrast, the ZPE of TS in the new model rises more for the ^{13}C isotope. Thus, despite the different signs in ΔZPE, in both cases the ^{12}C isotope is favored in terms of the reaction rate.

A.2 Details on LHCII Results

A.2.1 Other LHCII subunits

The HOMO → LUMO mixing of Chl and Vio was found to be strong in monomer A of the PDB entry 2BHW [132], when using CAM-B3LYP/6-31G(d) and the PCF. To see if the findings hold also for the other monomers, we calculated the most important spectra for the two other monomers B and C as well. The results are summed up in tab. A.4, together with monomer A.

Monomer	Model	Weak Band			
		λ_{max}/nm (eV)	f_I	% Chl	% Vio
A	3	605 (2.05)	0.9630	63.1	21.7
	3+PCF	609 (2.04)	1.9736	34.4	51.7
B	3	603 (2.06)	0.5968	72.7	11.7
	3+PCF	609 (2.04)	0.7715	60.9	19.7
C	3	599 (2.07)	0.9026	65.9	19.4
	3+PCF	605 (2.05)	1.0680	54.2	24.0
Monomer	Model	Strong Band			
		λ_{max}/nm (eV)	f_I	% Chl	% Vio
A	3	588 (2.11)	3.0949	16.3	75.6
	3+PCF	594 (2.09)	1.9747	35.0	44.1
B	3	578 (2.14)	3.3065	7.4	82.6
	3+PCF	589 (2.10)	3.1740	14.0	74.6
C	3	579 (2.14)	3.2113	15.0	77.5
	3+PCF	588 (2.11)	2.9772	18.5	56.9

Table A.4: Overview of the details of the CAM-B3LYP/6-31G(d) Chlorophyll b and Violaxanthin bands with strong HOMO → LUMO character for the coupled systems *in vacuo* and in point charge field (models 3 and 3+PCF) for all three LHCII monomers. Characters other than HOMO ↔ LUMO omitted for the analysis. The percentage of contribution of each character to a state is given by (coefficient)^2 × 200% (see tab. 5.2).

It can be seen while the mixing is strongest in monomer A, the changes upon introducing a point charge field are the same for all three monomers. The oscillator strength of the weaker $S_{0\to1}^{Chl}$ band is increased, and the HOMO → LUMO character in both states becomes more evenly distributed.

A.2.2 Effect of Supporting Amino Acids

An incomplete Mg^{2+} coordination sphere (omitting the Lys23, Tyr24 and Leu25 chain) has little effect on the spectrum of chlorophyll b. The corresponding spectra are displayed in fig. A.1.

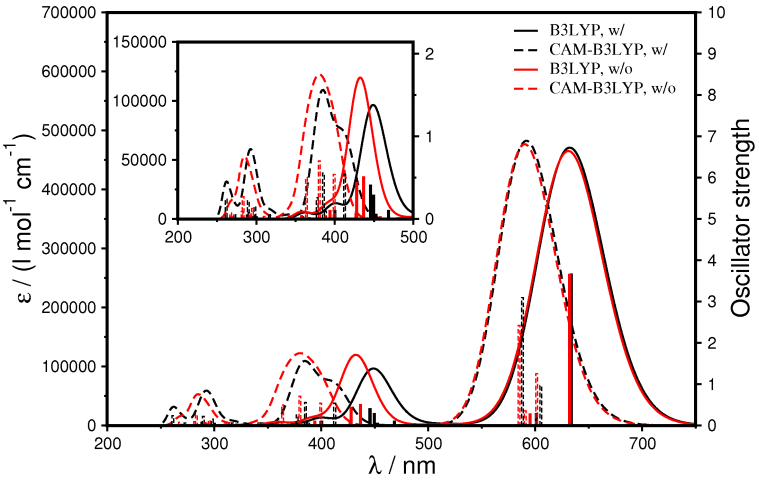

Figure A.1: Calculated 6-31G(d) spectra, 30 singlet states, of the coupled chromophores chlorophyll b and violaxanthin in crystal geometry (model 2). With (w/) and without (w/o) amino acid chain coordinated to the chlorophyll b Mg^{2+} ion. See text for details.

While there are slight blue shifts of energetically higher bands, the $S^{Chl}_{0\to1}$ transition remains unaffected. Since our analysis is centered on the $S^{Chl}_{0\to1}$ transition, we can consider the spectra to be equivalent (in case of using the same computational method).

A.2.3 Effect of the Phytyl Residue

The phytyl residue of chlorophyll b was included in all corresponding calculations presented in chapter 5. The changes occurring when not taking the phytyl group into account are shown in fig. A.2.

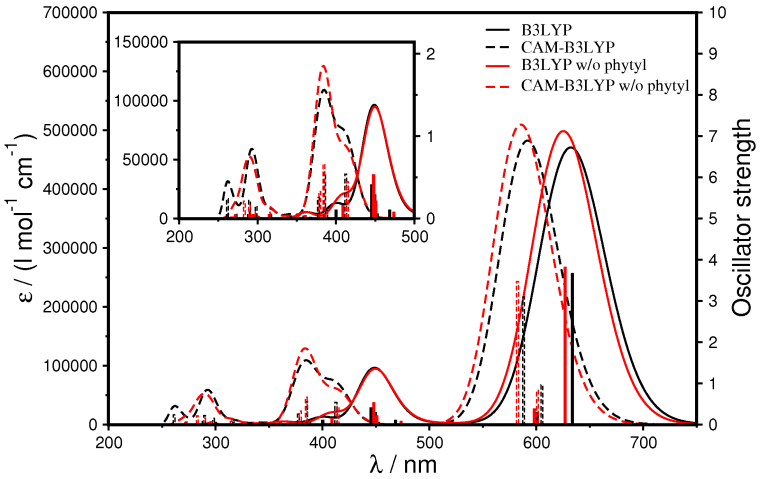

Figure A.2: Calculated 6-31G(d) spectra, 30 singlet states, of the coupled chromophores chlorophyll b and violaxanthin in crystal geometry (model 2). With (w/) and without (w/o) the phytyl group of the chlorophyll b. If omitted, the phytyl group was replaced by a hydrogen.

It can be seen that the phytyl group has a small effect on the position of the violaxanthin band. As such, the phytyl group might be a target for future investigations, if the coupling scheme proposed in chapter 5 is supported by other approaches. For now, it suffices to say that the phytyl residue should be included, as it affects the $S_{0\rightarrow1}$ transitions in our system.

A.2.4 Effect of His Charge

We chose the His residues to be singly protonated only for our main study. For comparison, we recalculated the CAM-B3LYP models for the three monomers with fully protonated His. The results are listed in tab. A.5.

Monomer	Model	Weak Band			
		λ_{max}/nm (eV)	f_I	% Chl	% Vio
A	2^{Chl}+PCF	597 (2.08)	0.1455	n/a	
	3+PCF	604 (2.05)	1.3208	47.5	33.6
B	2^{Chl}+PCF	602 (2.06)	0.1661	n/a	
	3+PCF	607 (2.04)	0.5272	63.8	10.5
C	2^{Chl}+PCF	597 (2.08)	0.1559	n/a	
	3+PCF	602 (2.06)	0.7010	64.2	17.2
Monomer	Model	Strong Band			
		λ_{max}/nm (eV)	f_I	% Chl	% Vio
A	2^{Vio}+PCF	583 (2.13)	4.2614	n/a	
	3+PCF	588 (2.11)	2.6860	22.8	58.8
B	2^{Vio}+PCF	575 (2.16)	4.2496	n/a	
	3+PCF	581 (2.13)	3.5120	8.0	76.7
C	2^{Vio}+PCF	574 (2.16)	4.3569	n/a	
	3+PCF	581 (2.13)	3.4093	11.7	79.1

Table A.5: Overview of the details of the CAM-B3LYP/6-31G(d) Chlorophyll b and Violaxanthin bands with strong HOMO → LUMO character for the isolated molecules and coupled systems, both in point charge field (models 2+PCF and 3+PCF) for all three LHCII monomers with all-positive His residues in the PCF. Characters other than HOMO ↔ LUMO omitted for the analysis. The percentage of contribution of each character to a state is given by (coefficient)^2 × 200% (see tab. 5.2).

In comparison to tab. A.4, the mixing of the HOMO → LUMO character is found to be weaker in all cases. Consequently, the effect of the His protonation seems to be systematic. If the mixing of the HOMO → LUMO transitions is indeed a way for light harvesting, the diminishing of mixing upon His protonation would make sense: Most His residues (6 of 9) are located at the luminal side of the LHCII, which experiences acidification upon excess light. Acidification leads to His protonation, which in turn might "decouple" the chromophores to stop light harvesting under excess light conditions.

A.2.5 Effect of Basis Set

To see the effect of the basis set on our calculated spectra, we conducted a basis set test with B3LYP and the coupled chromophore system (including the supporting amino acid chain). The results can be seen in fig. A.3.

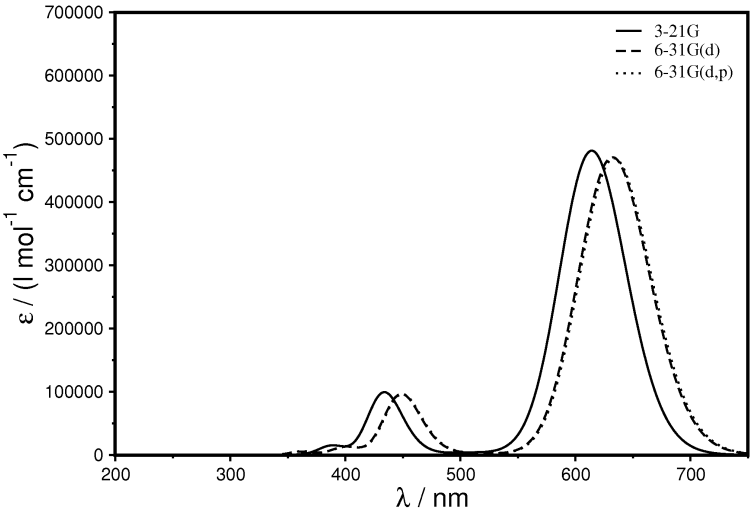

Figure A.3: Calculated B3LYP spectra, 30 singlet states, of the coupled chromophores chlorophyll b and violaxanthin in crystal geometry (model 3).

While the basis set 3-21G does show a high blue shift for all bands, all of them are still present. The spectra of 6-31G(d) and 6-31G(d,p) are virtually identical. Since the latter increases the computational effort significantly (approx. 160 hydrogens, 480 p-functions), we stick with the 6-31G(d) basis set for all calculations shown in chapter 5.

A.2.6 CIS results

As wave function based methods do not suffer from long-range deficiencies as DFT, we calculated éxemplary selected spectra using CIS. The results are depicted in fig. A.4.

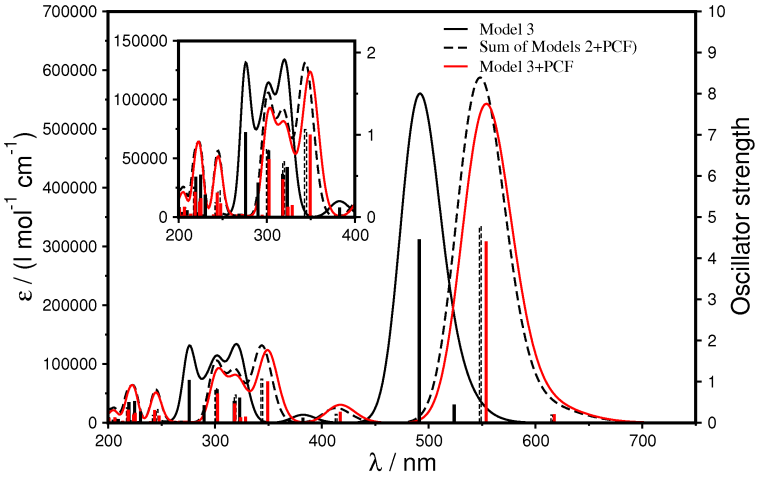

Figure A.4: Calculated CIS/6-31G(d) spectra, 30 singlet states, of selected models.

Is was found that the trend seen in tab. 5.1 is conserved throughout all models. The relative position of the CIS maxima in comparison to the DFT maxima of the different models is always at shorter wave lengths. Yet, we see that the qualitative effects seen in the DFT case hold for CIS, too: Introduction of the crystal geometry as well as the point charge field and the pairing of the chromophores all induce a red shift. While the strength of the shifts are mostly comparable to the DFT cases, the shift when introducing a point charge field is about 10 times stronger than in the DFT case.

We conclude that regardless of using a DFT based or wave function based method, the quality of our results stays the same (strong red shift of Vio, weak red shift of Chl). Note that the CIS case does not predict the Chl and Vio bands to be energetically close, which would forbid the mixing found in the DFT results.

Erklärung

Ich erkläre hiermit, dass ich diese Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe.

Jan Philipp Götze (Potsdam, den 15. September 2010)