Computational Studies of Photobiological Keto-Enol Reactions and Chromophores

Olle Falklöf

Department of Physics, Chemistry, and Biology (IFM)
Linköping University, SE-581 83 Linköping, Sweden

Linköping 2015

ISBN 978-91-7685-922-3
ISSN 0345-7524

Printed by LiU-Tryck 2015

Abstract

This thesis presents computational chemistry studies of keto-enol reactions and chromophores of photobiological significance.

The first part of the thesis is concerned with two protein-bound chromophores that, depending on the chemical conditions, can exist in a number of different ketonic and enolic forms. The first chromophore is astaxanthin, which occurs in the protein complex responsible for the deep-blue color of lobster carapace. By investigating how different forms of astaxanthin absorb UV-vis radiation of different wavelengths, a model is presented that explains the origin of the dramatic color change from deep-blue to red upon cooking of live lobsters.

The second chromophore is the oxyluciferin light emitter of fireflies, which is formed in the catalytic center of the enzyme firefly luciferase. To date, there is no consensus regarding which of the possible ketonic and enolic forms is the key contributor to the light emission. In the thesis, the intrinsic tendency of oxyluciferin to prefer one particular form over other possible forms is established through calculation of keto-enol and acid-base excited-state equilibrium constants in aqueous solution.

The second part of the thesis is concerned with two families of biological photoreceptors: the blue-light-absorbing LOV-domain proteins and the red-light-absorbing phytochromes. Based on the ambient light environment, these proteins regulate physiological and developmental processes by switching between inactive and active conformations. In both families, the conversion of the inactive into the active conformation is triggered by a chemical reaction of the respective chromophore.

The LOV-domain proteins bind a flavin chromophore and regulate processes such as chloroplast relocation and phototropism in plants. An important step in the activation of these photoreceptors is a singlet-triplet transition between two electronically excited states of the flavin chromophore. In the thesis, this transition is used as a prototype example for illustrating, for the first time, the ability of first-principles methods to calculate rate constants of inter-excited state phosphorescence events.

Phytochromes, in turn, bind bilin chromophores and are active in the regulation of processes like seed germination and flowering time in plants. Following two systematic studies identifying the best way to model the UV-vis absorption

and fluorescence spectra of these photoreceptors, it is demonstrated that steric interactions between the chromophore and the apoprotein play a decisive role for how phytochromes are activated by light.

Populärvetenskaplig sammanfattning

Solljus är en förutsättning för liv på jorden. Växter använder solljus som energikälla för att kunna bilda näring till sig själva och även syre till andra organismer. I ögonen hos människor och djur finns specialdesignade celler som omvandlar ljus till elektriska impulser, som sedan tolkas till en bild av syncentrat i hjärnan. Djur har utvecklat mekanismer för att med hjälp av ljus kunna kamouflera sig, locka till sig byten och kommunicera med varandra. Den gren av kemin som studerar kemiska processer som antingen initieras av eller ger upphov till ljus kallas för *fotokemi*. I denna avhandling har datorbaserad modellering använts för att skapa en bättre förståelse för de grundläggande kemiska mekanismer som styr olika ljuskänsliga proteiner. För att dessa proteiner ska fungera utnyttjar de en eller flera hjälpmolekyler som absorberar ljus. Dessa molekyler kallas för kromoforer och ger proteinerna deras karakteristiska färger.

Den första delen av avhandlingen behandlar kromoforer som kan anta en rad olika kemiska former med vitt skilda fotokemiska egenskaper. Det första exemplet som studeras är astaxantin som ger levande humrar deras karakteristiska mörkblå färg, men likaledes är orsak till att humrar blir röda när man kokar dem. Vad som ligger bakom denna dramatiska färgförändring har länge varit ett mysterium, men i avhandlingen presenteras en modell som på atomär nivå kan förklara detta fenomen. Vidare undersöks också vilken av de möjliga formerna av kromoforen oxyluciferin som ansvarar för eldflugornas lysande sken.

Den andra delen av avhandlingen behandlar frågeställningar rörande hur biologiska fotoreceptorer aktiveras. Biologiska fotoreceptorer är en grupp av proteiner som känner av och omsätter rådande ljusförhållanden till ett biologiskt gensvar. Speciellt studeras två olika familjer av biologiska fotoreceptorer. Den första familjen är de så kallade LOV-domänproteinerna, som exempelvis bidrar till att växter föredrar att växa i riktning mot solen och som aktiveras genom en mekanism som sker i flera kemiska steg. Vid ett av stegen sker en övergång från ett ljusupptagande elektroniskt tillstånd till ett reaktivt tillstånd vari en kemisk reaktion sedan inträffar. I avhandlingen används en nyutvecklad modelleringsmetod för att

avgöra exakt hur denna övergång sker. Den andra familjen kallas för fytokromer och reglerar bland annat växters vårblomning. I avhandlingen undersöks först vilken modelleringsmetod som är bäst lämpad för att bestämma färgen på det ljus som dessa proteiner tar upp och avger, och därefter studeras den grundläggande fotokemiska mekanism varigenom fytokromer aktiveras.

Acknowledgements

First of all I would like to thank my supervisor Bo Durbeej, without whom this thesis would not have been possible. Thank you Bo, for all your invaluable help, support and guidance during the past five years.

I would also like to thank my co-supervisor Patrick Norman for all his help and assistance, and for introducing me to relativistic quantum chemistry.

Thanks to Michele Cianci, Alfons Hädener, John Helliwell, Madeleine Helliwell, Andrew Regan and Ian Watt for collaboration on the astaxanthin project.

I am grateful to Lejla Kronbäck for taking care of all administrative issues and Magnus Boman for administrating my teaching.

All present and former colleagues in the Theoretical Chemistry group (formerly Computational Physics group) deserve great thanks for making my working days more pleasant. My special thanks go to Baswanth Oruganti and Chang-Feng Fang.

Ett stort tack går till mamma, pappa, Anneli och Anders för allt ert stöd och för all er hjälp under skrivandet av den här avhandlingen.

Avslutningsvis vill jag rikta ett stort tack till de som står mig allra närmast: Joanna, Julie och Sander, ni är ljuset i mitt liv!

Olle Falklöf

Linköping, November 2015

Contents

1	Introduction	3
2	Photochemical Modeling	7
2.1	Basic Quantum Chemistry	8
2.1.1	Hartree-Fock Theory	10
2.1.2	Electron Correlation	11
2.1.3	Density Functional Theory	12
2.2	Time-Dependent Density Functional Theory	15
2.3	Other Methods for Modeling Excited States	18
2.3.1	Configuration Interaction Methods	18
2.3.2	Coupled Cluster Methods	19
2.3.3	Multiconfigurational Methods	19
2.4	Relativistic Effects	19
2.4.1	The Dirac Equation	20
2.4.2	Approximate Relativistic Approaches	21
2.5	Environmental Effects	22
2.5.1	Polarizable Continuum Models	22
2.5.2	QM/MM Methods	25
2.6	Concluding Remarks	28
3	Photobiological Keto-Enol Reactions	29
3.1	Keto-Enol Tautomerism	29
3.2	Astaxanthin	32
3.2.1	Paper I	33
3.3	Oxyluciferin	38
3.3.1	Paper II	39

4	Biological Photoreceptors	43
4.1	Overview	43
4.2	LOV-Domain Proteins	46
4.2.1	Paper III	47
4.3	Phytochromes	50
4.3.1	Paper IV	53
4.3.2	Paper V	55
4.3.3	Paper VI	57
5	Concluding Remarks	61
6	Publications Included in the Thesis	75
	Paper I	77
	Paper II	119
	Paper III	155
	Paper IV	213
	Paper V	249
	Paper VI	277

Acronyms

4C Four-component

AXT Astaxanthin

BH Born-Haber

BLUF Blue-light sensors using flavin-adenine dinucleotide

BO Born-Oppenheimer

BV Biliverdin IX α

CASPT2 Complete active space with second-order perturbation theory

CASSCF Complete active space self-consistent field

CC Coupled cluster

CC2 Second-order approximate coupled cluster singles and doubles

CI Configuration interaction

CIS Configuration interaction singles

CIS(D) CIS with perturbation-theory treatment of doubles corrections

DFT Density functional theory

DrBphP *Deinococcus radiodurans* bacteriophytochrome

FMN Flavin mononucleotide

GGA Generalized gradient approximation

HF Hartree-Fock

HK Hohenberg-Kohn

KS Kohn-Sham

LDA Local density approximation

LF Lumiflavin

LR Linear response

MCSCF Multiconfigurational self-consistent field

MM Molecular mechanics

ONIOM Our N-layered integrated molecular orbital and molecular mechanics

OxyLH₂ Oxyluciferin

PCM Polarizable continuum model

PES Potential energy surface

Pfr Far-red-absorbing form of phytochromes

Pr Red-absorbing form of phytochromes

QM Quantum mechanics

QM/MM Quantum mechanics/molecular mechanics

SAC-CI Symmetry-adapted cluster configuration interaction

SS State specific

TD-DFT Time-dependent density functional theory

TF Thomas-Fermi

UVR8 Ultraviolet resistance locus 8

UV-vis Ultraviolet-visible

λ_{max} Absorption maximum

CHAPTER 1

Introduction

Sunlight is a prerequisite for life on Earth. Sunlight provides energy for plants to make carbohydrates and oxygen, which are used by themselves and other living organisms. The human eye contains light-sensitive pigments that generate electrochemical signals for the brain to interpret. Some animals have developed mechanisms that enable production of light inside their bodies such as the colorful jellyfish in the oceans and the glowing fireflies lighting up the night sky. These examples illustrate only a few of the numerous chemical and biological processes that involve light. In the field of photochemistry, scientists not only observe such processes, but also try to understand and explain the mechanisms behind them at molecular, atomic and even electronic levels.

The common denominator for all photochemical processes is that they involve electronically excited states. According to the International Union of Pure and Applied Chemistry (IUPAC),¹ an electronically excited state is “a state of an atom or molecular entity which has greater electronic energy than the ground state of the same entity.” The chemical properties of a molecule in an excited state is usually substantially different from those in the ground state. For example, a molecule that is weakly acidic in the ground state may become strongly acidic in an excited state, and a chemical bond that has double-bond character in the ground state may obtain single-bond character in an excited state. Thus, chemistry that is not possible in the ground state may well occur in an excited state.

Since molecules always strive to reach their lowest possible energy state, a photochemical reaction is always facing a competition with a number of different relaxation processes. The relaxation from one electronic state to another can occur either through light emission (a so-called radiative process) or dissipation of excess energy as heat into the environment (a so-called non-radiative process). Radiative processes include fluorescence and phosphorescence, and can often be observed using UV-vis (ultraviolet-visible) spectroscopy. Non-radiative processes,

in turn, include internal conversion and intersystem crossing. While fluorescence and internal conversion take place between states with equal spin multiplicities (e.g., singlet-singlet and triplet-triplet transitions), phosphorescence and intersystem crossing involve states of different spin multiplicities (e.g., singlet-triplet and triplet-singlet transitions). Since none of these processes changes the chemical identity of the system at hand, they are usually referred to as photophysical processes. Another example of a photophysical process is light absorption, which raises the energy of the system and results in an excited state being formed. All these photophysical processes are illustrated in Fig. 1.1.

Figure 1.1. Examples of photophysical processes. Here, absorption (A), fluorescence (F) and internal conversion (IC) occur between singlet states, whereas intersystem crossing (ISC) and phosphorescence (P) involve singlet and triplet states.

During the past decades, much effort has been put into the development of accurate and cost-effective computational methods for studying photophysical and photochemical processes. In the early 50's, Pariser, Parr and Pople implemented one of the first methods to compute and predict UV-vis spectra of unsaturated molecules.^{2,3} Although their method offered an accurate description of $\pi \rightarrow \pi^*$ excitations within the valence shell, it was not clear how to treat other types of transitions like Rydberg and $n \rightarrow \pi^*$ transitions. For a more general treatment of excited states, new computational methods were needed.

In the 80's, multiconfigurational self-consistent field methods, such as the complete active space self-consistent field approach by Roos and co-workers,⁴ became available for calculating, in a balanced way, both ground and excited states. With these methods, it was possible to follow the molecular motions in the excited states utilizing analytical gradient techniques. This opened up the door for computational studies of photochemical reactions. In 1990, Robb and co-workers,⁵ for example, used this approach to study the photoinduced cycloaddition of two ethylene molecules, and located a conical intersection (surface crossing) between the lowest excited (singlet) state and the ground state.

While multiconfigurational self-consistent field methods in many cases give

accurate results, they are computationally expensive and therefore limited in applicability to rather small systems. For modeling large molecular systems, on the other hand, cheaper methods like configuration interaction singles typically perform quite poorly.⁶ A breakthrough in the modeling of excited states of large molecular systems was the advent of time-dependent density functional theory in the late 90's.⁷⁻¹¹ This method has similar computational requirements as the configuration interaction singles approach but performs markedly better. Furthermore, just like its parent theory (density functional theory), time-dependent density functional theory is a formally exact theory.

While time-dependent density functional theory is the standard choice of methodology for modeling large molecular systems containing up to, say, hundreds of atoms, it is still not practical to perform such calculations on systems containing thousands of atoms, such as proteins. In many proteins, however, the photophysical and photochemical processes are relatively localized to the chromophore. At the beginning of the 21st century, the so-called hybrid quantum mechanics/molecular mechanics approach, which combines the accuracy of quantum chemical methods (needed to describe the chromophore) with the computational efficiency of molecular mechanics methods (needed to describe the full system), started to find widespread use in photochemical applications.¹² Among the studies reported in the literature during that time, one may highlight the study by Olivucci and co-workers investigating the excited-state dynamics of the retinal chromophore present in rhodopsin.¹³ Although significant methodological improvements have been made during the past decades, it is still a very challenging task to describe photochemical, as well as photophysical, phenomena. This, together with all the fascinating applications, makes photochemistry an exciting field of research.

The present thesis deals with computational studies of photophysical and photochemical phenomena in biological systems. In the first part of the thesis, quantum chemical calculations are performed to study two-protein bound chromophores that, depending on the chemical conditions, can exist in a number of different ketonic and enolic forms. The first chromophore is astaxanthin, which is bound to a protein complex responsible for the deep-blue color of live lobsters. By investigating how the absorption maxima depends on the chemical identity of the astaxanthin chromophore, a model is presented that explains why lobsters turn red upon cooking.

The second chromophore is the oxyluciferin chromophore that is responsible for the fluorescence of fireflies, and is formed in a reaction catalyzed by the enzyme firefly luciferase. Today, there is no consensus regarding which of the different ketonic and enolic forms is the main contributor to the observed fluorescence. In the thesis, the intrinsic tendency of oxyluciferin to favor one particular form over other possible forms is predicted by calculating keto-enol and acid-base excited-state equilibrium constants in aqueous solution.

In the second part of the thesis, quantum chemical and hybrid quantum mechanics/molecular mechanics calculations are performed to study two families of biological photoreceptors: the blue-light-sensing LOV-domain proteins and the red-light-sensing phytochromes. These proteins exist in inactive and active forms, where the relative concentrations of the two forms govern physiological and devel-

opmental processes in the host organism. In both families, the conversion from the inactive to the active form begins with a photochemical reaction of the respective chromophore.

The LOV-domain proteins contain a flavin chromophore and are active in the regulation of processes like chloroplast relocation and phototropism in plants. A critical step in the activation of LOV-domain proteins is a transition from an electronically excited singlet state to an electronically excited triplet state. In the thesis, this transition is used as an example to illustrate, for the first time, that first-principles calculations can be performed to obtain rate constants for phosphorescence events occurring between two different excited states.

Phytochromes, in turn, contain bilin chromophores and regulate processes such as seed germination and flowering time in plants. Following two systematic studies aimed at identifying the optimal way to use quantum chemical methods to model photophysical processes of these photoreceptors, it is demonstrated that steric effects is the key mechanism by which phytochromes control the photochemical reactivity of their bilin chromophores.

The materials presented in this thesis, which have been revised and extended from a previous licentiate thesis written by the author,¹⁴ is organized into five chapters. In Chapter 2, the theory underlying the methods used are introduced. Chapters 3 and 4 present the background to the different projects and summarize the key results of the respective papers. Chapter 5 gives some concluding remarks. Finally, the papers, together with their supporting informations, are included as appendices.

CHAPTER 2

Photochemical Modeling

A chemical reaction can be classified as either a thermal or a photochemical reaction. While thermal reactions take place exclusively in the electronic ground state, photochemical reactions also involve one or several electronically excited states. In order to study photochemical reactions by means of theoretical methods, the computational chemist therefore needs to compute, in a balanced way, relevant parts of both ground and excited-state potential energy surfaces (this concept is explained below). Although the calculation of a ground-state potential energy surface and its key features (e.g., minima and transition states) is a demanding task, the treatment of an excited-state potential energy surface is often even more challenging. In fact, compared to a ground-state potential energy surface, excited-state potential energy surfaces usually also involve crossings with and couplings to other excited states. To make things even more complicated, there are also many different types of excited states, each with its own requirements on the methodology.

A photochemical process usually starts with a photoexcitation from the ground state to an absorbing excited state. To predict which of the possible excited states is the absorbing one, one can calculate the transition energies and transition dipole moments to obtain oscillator strengths, which yield information on the intensities of the different transitions. Following photoexcitation, the system will start to move from the so-called Franck-Condon region toward the nearest local minimum on the excited-state potential energy surface of the absorbing state, which in many cases corresponds to the initial stages of a chemical reaction. In other cases, however, the system will not reach the minimum before it decays to a lower-lying electronic state. By performing calculations to explore the excited-state potential energy surface, one can obtain rate constants for chemical reactions and compare these values with rate constants for the competing relaxation processes. If a chemical reaction is faster than the competing relaxation processes, then the reaction will occur before the system relaxes into a lower-lying electronic state.

Otherwise, the system will lose energy by either emitting light or dissipating excess energy as heat into the environment.

Closely related to the treatment of photochemical processes is the modeling of photophysical properties such as UV-vis absorption and fluorescence spectra. To calculate such spectra, one first uses some optimization algorithm to locate some relevant points on the ground and excited-state potential energy surfaces. Thereafter, energies and oscillator strengths for a large number of transitions are computed to provide the ranges and the forms of the absorption and fluorescence energy bands. For such calculations, it is important that all transitions that contribute to the bands are described with the same accuracy. However, since all methods inevitably describe some transitions better than others, it can be a difficult task to reproduce the shape of experimentally recorded spectra.

This chapter will present the general theory underlying the methods used in the thesis for describing photophysical and photochemical processes. All equations are given in atomic units (a.u.), which means that $m = \hbar = e = 4\pi\epsilon_0 = 1$ (a.u.), where m is the electron mass, $\hbar$ is Planck's constant divided by 2π , e is the elementary charge, and ϵ_0 is the vacuum permittivity. Furthermore, to distinguish between various types of molecular orbitals, indices i and j refer to occupied orbitals, indices a and b to virtual orbitals, and k , l , m and n to general orbitals throughout the presentation.

2.1 Basic Quantum Chemistry

The starting point for a non-relativistic quantum mechanical treatment of a molecular system consisting of M nuclei and N electrons is the time-dependent Schrödinger equation (given in atomic units)

$$\hat{H}\Psi = i\frac{\partial\Psi}{\partial t}, \quad (2.1)$$

where Ψ is the wave function describing the state of the system. The Hamiltonian operator $\hat{H}$ is in this framework given by

$$\hat{H} = \hat{T}_e + \hat{T}_n + \hat{V}_{ee} + \hat{V}_{nn} + \hat{V}_{en}, \quad (2.2)$$

where $\hat{T}_e$ is the kinetic energy operator of the electrons, $\hat{T}_n$ is the kinetic energy operator of the nuclei, and $\hat{V}_{ee}$, $\hat{V}_{nn}$ and $\hat{V}_{en}$ are the electron-electron, nucleus-nucleus and electron-nucleus interaction operators, respectively. If the Hamiltonian is time-independent, the solution of the time-dependent Schrödinger equation takes the form

$$\Psi(\mathbf{R}, \mathbf{r}, t) = \psi(\mathbf{R}, \mathbf{r})e^{-iEt}, \quad (2.3)$$

where the wave function $\psi(\mathbf{R}, \mathbf{r})$ describes a stationary state with energy E , and $\mathbf{r}$ and $\mathbf{R}$ represent all spatial coordinates of the electrons and nuclei, respectively. By inserting Eq. 2.3 into the time-dependent Schrödinger equation (Eq. 2.1), the time-independent Schrödinger equation,

$$\hat{H}\psi(\mathbf{R}, \mathbf{r}) = E\psi(\mathbf{R}, \mathbf{r}), \quad (2.4)$$

is obtained. This equation is central in quantum chemistry.

Expanding $\psi(\mathbf{R}, \mathbf{r})$ in an orthonormal set of electronic wave functions, $\{\psi_{ei}(\mathbf{r}; \mathbf{R})\}_{i=1}^{\infty}$, the exact solution of the time-independent Schrödinger equation (Eq. 2.4) can then be written as

$$\psi(\mathbf{R}, \mathbf{r}) = \sum_{i=1}^{\infty} \psi_{ni}(\mathbf{R}) \psi_{ei}(\mathbf{r}; \mathbf{R}). \quad (2.5)$$

Here, the nuclear wave functions $\{\psi_{ni}(\mathbf{R})\}_{i=1}^{\infty}$ are considered to be expansion coefficients, and $\psi_{ei}(\mathbf{r}; \mathbf{R})$ depends *explicitly* on the electronic coordinates and *parametrically* on the nuclear coordinates. Multiplying Eq. 2.5 to the left with the electronic wave function of state j and integrating over the electronic coordinates gives

$$(\hat{T}_n + E_{ej} + \langle \psi_{ej} | \hat{T}_n | \psi_{ej} \rangle) \psi_{nj} + \Lambda = E \psi_{nj}, \quad (2.6)$$

where the electronic energy, $E_{ej} = E_{ej}(\mathbf{R})$, is a function of the nuclear geometry and Λ represents the so-called non-adiabatic coupling elements.

In the adiabatic approximation the electronic wave function is assumed to be restricted to one specific electronic state. This means that all non-adiabatic coupling elements in Eq. 2.6 can be neglected and a simplified equation

$$(\hat{T}_n + E_{ej} + \langle \psi_{ej} | \hat{T}_n | \psi_{ej} \rangle) \psi_{nj} = E \psi_{nj} \quad (2.7)$$

for the total energy E is obtained. To further simplify the description, the $\langle \psi_{ej} | \hat{T}_n | \psi_{ej} \rangle$ term can be neglected, because of the large difference in mass between electrons and nuclei. This approximation, commonly known as the Born-Oppenheimer (BO) approximation, is usually very appropriate, provided that the electronic wave function is a slowly varying function of the nuclear coordinates. However, for conical intersections or for other situations where two different states become (nearly) degenerate, which frequently occur in photochemical processes, both the adiabatic and BO approximations break down. The resulting equation after the adiabatic and BO approximations have been introduced is

$$(\hat{T}_n + E_{ej}) \psi_{nj} = E \psi_{nj}, \quad (2.8)$$

where $E_{ej}(\mathbf{R})$ is a solution of the electronic Schrödinger equation

$$\hat{H}_e \psi_{ej}(\mathbf{r}; \mathbf{R}) = E_{ej}(\mathbf{R}) \psi_{ej}(\mathbf{r}; \mathbf{R}). \quad (2.9)$$

If ψ_{ej} describes the lowest (in energy) state, then this is the wave function for the electronic ground state. All other solutions to the electronic Schrödinger equation then correspond to electronically excited states.

Another consequence of the large difference in mass between electrons and nuclei is that the nuclear component of the wave function is spatially more localized than the electronic component and, accordingly, the electronic motion is assumed to take place in the field of fixed nuclei. By calculating the electronic energy at

different nuclear geometries, a so-called potential energy surface (PES) is obtained. Since the total nucleus-nucleus repulsion energy does not depend on the electronic coordinates, this energy term can be added to the other energy terms once the electronic problem has been solved. Thus, the electronic Hamiltonian can be written as $\hat{H}_e = \hat{T}_e + \hat{V}_{en} + \hat{V}_{ee}$.

In applications, the Schrödinger equation is solved by dividing the calculations into two steps. First, the electronic Schrödinger equation is solved for a fixed nuclear geometry, which is the step that this thesis focuses on. Then, the resulting energy E_{ej} is used for solving the nuclear problem. To simplify the notation, the electronic Hamiltonian, wave functions and energies are hereafter denoted as $\hat{H}$, ψ and E , respectively. Thereby, Eq. 2.9 reads

$$\hat{H}\psi = E\psi. \quad (2.10)$$

Unfortunately, analytic solutions of this equation can only be obtained for one-electron systems. For other systems, further approximations are required.

2.1.1 Hartree-Fock Theory

The Hartree-Fock (HF) method is one of the simplest methods for describing electron-electron interactions, and is also the starting point for more advanced electronic-structure methods. In this method, the electron-electron interactions are treated in a mean-field fashion, which means that every single electron interacts with the average field generated by all the other electrons. Approximating the antisymmetric N -electron wave function in Eq. 2.10 by a single Slater determinant composed of orthonormal spin orbitals $\phi_i(\mathbf{x}_j)$,

$$\psi_{\text{HF}}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \dots) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_1(\mathbf{x}_1) & \phi_1(\mathbf{x}_2) & \dots & \phi_1(\mathbf{x}_N) \\ \phi_2(\mathbf{x}_1) & \phi_2(\mathbf{x}_2) & & \vdots \\ \vdots & & \ddots & \vdots \\ \phi_N(\mathbf{x}_1) & \dots & \dots & \phi_N(\mathbf{x}_N) \end{vmatrix}, \quad (2.11)$$

the HF equations are obtained as

$$\hat{f}(\mathbf{x}_j)\phi_i(\mathbf{x}_j) = \epsilon_i\phi_i(\mathbf{x}_j) \quad (2.12)$$

by applying the variational principle with the constraint that the spin orbitals remain orthonormal. The variational principle states that for any trial wave function ψ_t

$$E_0 \leq \frac{\langle \psi_t | \hat{H} | \psi_t \rangle}{\langle \psi_t | \psi_t \rangle}, \quad (2.13)$$

where E_0 is the exact non-relativistic electronic ground-state energy. In Eq. 2.12, ϵ_i is the energy corresponding to the $\phi_i(\mathbf{x}_j)$ spin orbital, where the 4-dimensional coordinate $\mathbf{x}_j$ represents both spatial ($\mathbf{r}_j$) and spin (σ_j) variables. The Fock operator,

$$\hat{f}(\mathbf{x}_j) = -\frac{1}{2}\nabla_j^2 + \hat{v}_{\text{HF}}(\mathbf{x}_j), \quad (2.14)$$

is an *effective* one-electron operator that includes an effective one-electron potential operator $\hat{v}_{\text{HF}}(\mathbf{x}_j) = \hat{v}_{\text{en}}(\mathbf{x}_j) + \hat{J}(\mathbf{x}_j) + \hat{K}(\mathbf{x}_j)$. $\hat{v}_{\text{HF}}(\mathbf{x}_j)$ describes the interactions between an electron with all nuclei and all other electrons. While the Hartree (Coulomb) operator $\hat{J}(\mathbf{x}_j)$ is a local operator and corresponds to the classical electrostatic repulsion, the exchange operator $\hat{K}(\mathbf{x}_j)$ is a non-local operator and depends on the spin orbitals throughout space. The exchange operator arises from the antisymmetry requirement of the wave function and separate electrons of parallel spins (Fermi correlation), and is purely quantum mechanical in nature.

Expanding the spin orbitals in a set of basis functions (a so-called basis set) and solving the resulting equations (the so-called Roothaan-Hall equations), the ground-state energy is given by

$$E_{\text{HF}} = \sum_{i=1}^N \epsilon_i - \frac{1}{2} V_{\text{ee}} + V_{\text{nn}} = \sum_{i=1}^N \epsilon_i - \frac{1}{2} (J - K) + V_{\text{nn}}. \quad (2.15)$$

In this expression, the second term, which has been divided into Coulomb (J) and exchange (K) parts, corrects for the double counting of the electron-electron repulsion that arises by summing the orbital energies.

2.1.2 Electron Correlation

While HF often captures a large part of the total energy (with a large basis set),¹⁵ the remaining energy corresponding to correlated electronic motions is often of importance for the description chemical bonding. In a given basis set, a standard way of defining the correlation energy E_{corr} is

$$E_{\text{corr}} = E_0 - E_{\text{HF}}, \quad (2.16)$$

where E_{HF} is the HF energy.

Electron correlation effects may be divided into short-range dynamical and long-range non-dynamical (static) parts. Dynamical correlation arises from the correlated movements of electrons. A prototypical example is the movements of the two 1s electrons in the ground-state of the helium atom. Examples of wave function (*ab initio*) methods that have been developed to treat dynamical correlation include configuration interaction, coupled cluster and many-body perturbation theory methods. Non-dynamical correlation, on the other hand, arises from near-degeneracies and typically comes into play when chemical bonds dissociate or when different electronic states lie close in energy. In order to capture non-dynamical correlation effects, multiconfigurational methods are required such as the complete active space self-consistent field method.⁴ If both dynamical and non-dynamical correlation effects are at play, the dynamical correlation energy can be added on top of a complete active space self-consistent field energy using second-order perturbation theory.¹⁶

Unfortunately, in standard implementations correlated wave function methods are not viable for routine calculations on large molecular systems (hundreds of atoms), as they often scale very poorly (with respect to the number of basis functions) and require a large number of basis functions for accurate results. An

alternative way of treating the electron correlation problem is given by density functional theory, which will be described in the next section.

2.1.3 Density Functional Theory

Density functional theory (DFT) offers a route for carrying out correlated calculations at a relatively small computational cost. The central quantity in DFT is the electron density $\rho(\mathbf{r}_1)$, which is related to the many-electron wave function through

$$\rho(\mathbf{r}_1) = N \int |\psi(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)|^2 d\sigma_1 d\mathbf{x}_2 \dots d\mathbf{x}_n, \quad (2.17)$$

where N is number of electrons that follows straightforwardly from the electron density as

$$N = \int \rho(\mathbf{r}) d\mathbf{r}. \quad (2.18)$$

For a molecular system consisting of N electrons subject to an external potential $v_{\text{ext}}(\mathbf{r})$, due to M nuclei, the many-electron Hamiltonian is given by

$$\hat{H} = - \sum_{i=1}^N \frac{1}{2} \nabla_i^2 + \sum_{i=1}^N v_{\text{ext}}(\mathbf{r}_i) + \sum_{i < j} \frac{1}{r_{ij}}. \quad (2.19)$$

In a pioneering work by Hohenberg and Kohn,¹⁷ they proved that $v_{\text{ext}}(\mathbf{r})$ is uniquely defined (by up to an additive constant) by the electron density. This is the first Hohenberg-Kohn (HK) theorem.¹⁷ The consequence of this theorem is that the Hamiltonian can be fully recreated from the electron density ($\hat{H} = \hat{H}[\rho(\mathbf{r})]$), and that the electron density determines the wave function and, thereby, all properties of the system at hand. In particular, the exact electronic energy of a system can be written as

$$E[\rho] = T[\rho] + V_{\text{ee}}[\rho] + V_{\text{ext}}[\rho] = F[\rho] + \int \rho(\mathbf{r}) v_{\text{ext}}(\mathbf{r}) d\mathbf{r}, \quad (2.20)$$

where $T[\rho]$ is kinetic energy functional and $F[\rho]$ is a (unknown) universal functional of ρ

$$F[\rho] = T[\rho] + V_{\text{ee}}[\rho] = \langle \psi | \hat{T} + \hat{V}_{\text{ee}} | \psi \rangle. \quad (2.21)$$

Hohenberg and Kohn also showed that for $v_{\text{ext}}(\mathbf{r})$, the exact ground-state electron density ρ_0 , derived from an N -electron antisymmetric wave function, can be obtained by variationally minimizing the energy, $E_0 = E[\rho_0]$. This is the second HK theorem.¹⁷

The Kohn-Sham Equations

Thomas-Fermi (TF) theory is considered to be the original version of density functional theory. In this theory, the kinetic energy $T[\rho]$ is given by a local functional of

the electron density, which is exact for a uniform gas of non-interacting electrons. While the TF functional reasonably reproduces atomic energies, it is unable to describe important chemical phenomena such as atomic shell structure and covalent bonds.¹⁸ The main flaw in TF theory is the way electron dynamics is incorporated into the functional, which causes a large error in the kinetic energy. One way to improve the TF model is to include (non-local) kinetic energy terms such as the von Weizsäcker correction.¹⁹ Although this inclusion improves the treatment of the kinetic energy, these orbital-free functionals are still inferior to HF.

Instead of explicitly deriving an expression of the kinetic energy in terms of ρ , Kohn and Sham²⁰ introduced a practical scheme to calculate the electronic energy using molecular orbitals. In this scheme, Kohn and Sham proposed that a real system of interacting electrons moving in an *external* potential $v_{\text{ext}}(\mathbf{r}_i)$ can be replaced by a fictitious system of non-interacting electrons moving in an *effective* potential $v_{\text{eff}}(\mathbf{r}_i)$ constructed in such a way that the density of the fictitious system is identical to that of the real system. For such a system of non-interacting electrons, the electronic Hamiltonian is simply

$$\hat{H} = \sum_{i=1}^N \left[-\frac{1}{2} \nabla_i^2 + v_{\text{eff}}(\mathbf{r}_i) \right] \quad (2.22)$$

and the exact ground-state wave function is given by a single Slater determinant formed from the orbitals corresponding to the N lowest (in energy) solutions of the Kohn-Sham (KS) equations

$$\left(-\frac{1}{2} \nabla_i^2 + v_{\text{eff}}(\mathbf{r}_i) \right) \phi_i(\mathbf{x}_i) = \epsilon_i \phi_i(\mathbf{x}_i). \quad (2.23)$$

The key advantage of the KS approach is that the complicated (real) kinetic energy functional $T[\rho]$ can be divided into a large part $T_S[\rho]$, which can be calculated exactly, and a small correction term $T[\rho] - T_S[\rho]$. The universal functional, $F[\rho]$, in the KS scheme is thus defined as

$$F[\rho] = T_S[\rho] + J[\rho] + E_{\text{xc}}[\rho], \quad (2.24)$$

where $T_S[\rho]$ is the exact kinetic energy of the system of non-interacting electrons given by

$$T_S[\rho] = \sum_{i=1}^N \langle \phi_i | -\frac{1}{2} \nabla_i^2 | \phi_i \rangle, \quad (2.25)$$

and $J[\rho]$ is the classical Hartree (Coulomb) repulsion expressed as

$$J[\rho] = \frac{1}{2} \int \frac{\rho(\mathbf{r}_1)\rho(\mathbf{r}_2)}{r_{12}} d\mathbf{r}_1 d\mathbf{r}_2. \quad (2.26)$$

The third term in Eq. 2.24 ($E_{\text{xc}}[\rho]$) is the *exchange-correlation functional* of the real system, which is defined as

$$E_{\text{xc}}[\rho] = T[\rho] - T_S[\rho] + V_{\text{ee}}[\rho] - J[\rho]. \quad (2.27)$$

Minimizing the exact energy of the real (interacting) system

$$E[\rho] = \int \rho(\mathbf{r})v_{\text{ext}}(\mathbf{r})d\mathbf{r} + T_{\text{S}}[\rho] + J[\rho] + E_{\text{xc}}[\rho], \quad (2.28)$$

subject to fixed number of electrons, gives the effective potential of the non-interacting system that generates the same electron density as that of the real system. This potential can be written as

$$v_{\text{eff}} = v_{\text{ext}} + \frac{\delta J[\rho]}{\delta \rho(\mathbf{r})} + \frac{\delta E_{\text{xc}}[\rho]}{\delta \rho(\mathbf{r})}. \quad (2.29)$$

Solving the KS equations using this potential gives the exact electronic ground-state energy of the system within the BO approximation. Unfortunately, since the exact analytic form of the exchange-correlation functional, $E_{\text{xc}}[\rho]$, is unknown for an arbitrary density, this functional has to be approximated. On the other hand, since the exchange-correlation energy is much smaller than the other energy terms, this is a significant improvement over orbital-free DFT such as TF theory. In the next section, some general ideas of the approximations to $E_{\text{xc}}[\rho]$ will be discussed.

Exchange-Correlation Functionals

During the past years, a plethora of different exchange-correlation functionals have been developed based on theoretical arguments and/or by fitting a number of parameters to experimental or high-level *ab initio* data.

The functionals can be divided into several classes.²¹ In the local density approximation (LDA), the functionals are based on the uniform electron gas model. Since LDA functionals significantly overestimate binding energies, however, they are not of widespread use in quantum chemistry. This approximation can be improved by adding energy terms that depend on the gradient of the electron density. This is referred to as the generalized gradient approximation (GGA). The next natural step is to introduce higher order derivatives into the functional such as the laplacian of the electron density. This yields the meta-GGA functionals. Functionals based on the LDA, GGA and meta-GGA approximations are normally referred to as pure functionals.

In order to further improve the energies, a fraction of HF (exact) exchange can be included in the functionals. These so-called hybrid functionals can be classified as global or range-separated hybrid functionals. While a global hybrid functional contains a fixed amount of HF exchange, the fraction of HF exchange included in a range-separated hybrid functional depends on the interelectronic distance. Since only the range-separated hybrid functionals exhibit the correct $-1/r$ asymptotic behavior, they usually give a much better description of Rydberg and charge-transfer states than other functionals.

The complexity of the exchange-correlation functional can be increased by including additional ingredients into the functional such as variables constructed from unoccupied KS orbitals. However, although the inclusion of additional variables into the functional allows for more flexibility, there is no systematic way to improve the accuracy of the results, which means that the performance of

exchange-correlation functionals have to be carefully assessed through benchmark calculations before they are applied to new chemical problems.

The DFT formalism based on the HK theorems¹⁷ can in principle be generalized to include the lowest state of a given space-spin symmetry, whereby one performs separate calculations on the ground and excited states to obtain excitation energies (the so-called Δ SCF approach). Although the Δ SCF approach can be useful if the excited state of interest has a different symmetry than the ground state, this methodology is of limited use for asymmetric molecules. For such systems, more general theoretical methods are necessary. One such method is time-dependent density functional theory,^{7–11} which will be described in the next section.

2.2 Time-Dependent Density Functional Theory

In 1984, Runge and Gross⁷ laid a firm foundation for time-dependent density functional theory (TD-DFT) by extending the basic ideas from static DFT into the treatment of excitation energies and general time-dependent phenomena. In their pioneering work, they proved that for a given initial state, the time-dependent charge density $\rho(\mathbf{r}, t)$ determines the external potential by up to an additive function of time. They also derived three practical schemes for calculating this density.

For an interacting system of electrons subject to a time-dependent external potential

$$v_{\text{ext}}(\mathbf{r}, t) = v_0(\mathbf{r}) + v_1(\mathbf{r}, t) \quad (2.30)$$

consisting of a static part $v_0(\mathbf{r})$ (typically the nuclear Coulomb potential) and a time-dependent perturbation $v_1(\mathbf{r}, t)$, the time-dependent KS equations for the non-interacting system take the form

$$\left[-\frac{1}{2}\nabla^2 + v_{\text{eff}}(\mathbf{r}, t)\right]\phi(\mathbf{r}, t) = i\frac{\partial}{\partial t}\phi(\mathbf{r}, t), \quad (2.31)$$

where the time-dependent effective potential $v_{\text{eff}}(\mathbf{r}, t)$ can be written as

$$v_{\text{eff}}(\mathbf{r}, t) = v_{\text{ext}}(\mathbf{r}, t) + \int \frac{\rho(\mathbf{r}', t)}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' + \frac{\delta A_{\text{xc}}[\rho]}{\delta \rho(\mathbf{r}, t)}. \quad (2.32)$$

Here, A_{xc} is the time-dependent exchange-correlation functional, which is the analogue of E_{xc} of the parent DFT approach. Normally, the adiabatic approximation

$$v_{\text{xc}}[\rho](\mathbf{r}, t) = \frac{\delta A_{\text{xc}}[\rho]}{\delta \rho(\mathbf{r}, t)} \approx \frac{\delta E_{\text{xc}}[\rho_t]}{\delta \rho_t(\mathbf{r})} = v_{\text{xc}}[\rho_t](\mathbf{r}) \quad (2.33)$$

is introduced. In this approximation, A_{xc} is assumed to be local in time, which makes it possible to use the exchange-correlation functionals of static DFT. The adiabatic approximation seems to work best for low-lying excited states.¹⁰

If the perturbation $v_1(\mathbf{r}, t)$ is turned on adiabatically, the first-order density response $\rho^{(1)}$ of the interacting system, initially in its ground state, is described

by the so-called linear response function $\chi_{\text{KS}}(t, t', \mathbf{r}, \mathbf{r}')$ of the non-interacting system.⁸ This takes the form

$$\begin{aligned} \rho^{(1)}(\mathbf{r}, t) &= \int \chi_{\text{KS}}(t, t', \mathbf{r}, \mathbf{r}') \left[v_1(\mathbf{r}', t') + \int \frac{\rho^{(1)}(\mathbf{r}'', t')}{|\mathbf{r}' - \mathbf{r}''|} d\mathbf{r}'' \right. \\ &\quad \left. + \int \frac{\delta^2 E_{\text{xc}}}{\delta \rho(\mathbf{r}') \delta \rho(\mathbf{r}'')} \rho^{(1)}(\mathbf{r}'', t') d\mathbf{r}'' \right] d\mathbf{r}' dt'. \end{aligned} \quad (2.34)$$

Since $\rho^{(1)}$ depends of $\rho^{(1)}$ itself, Eq. 2.34 has to be solved self-consistently with an iterative scheme.

Taking the Fourier transform with respect to time, one can transform Eq. 2.34 into the exact frequency-dependent linear density response within the adiabatic approximation⁸

$$\begin{aligned} \rho^{(1)}(\mathbf{r}, \omega) &= \int \chi_{\text{KS}}(\omega, \mathbf{r}, \mathbf{r}') \left[v_1(\mathbf{r}', \omega) + \int \frac{\rho^{(1)}(\mathbf{r}'', \omega)}{|\mathbf{r}' - \mathbf{r}''|} d\mathbf{r}'' \right. \\ &\quad \left. + \int f_{\text{xc}}(\mathbf{r}', \mathbf{r}'') \rho^{(1)}(\mathbf{r}'', \omega) d\mathbf{r}'' \right] d\mathbf{r}', \end{aligned} \quad (2.35)$$

where the exchange-correlation kernel

$$f_{\text{xc}}(\mathbf{r}', \mathbf{r}'') = \frac{\delta^2 E_{\text{xc}}}{\delta \rho(\mathbf{r}') \delta \rho(\mathbf{r}'')} \quad (2.36)$$

has been introduced, and $\chi_{\text{KS}}(\omega, \mathbf{r}, \mathbf{r}')$ can explicitly be expressed in terms of the KS orbitals as

$$\chi_{\text{KS}}(\omega, \mathbf{r}, \mathbf{r}') = \sum_{i,a} \left[\frac{\phi_i^*(\mathbf{r}) \phi_a(\mathbf{r}) \phi_i(\mathbf{r}') \phi_a^*(\mathbf{r}')}{\omega - (\epsilon_a - \epsilon_i)} - \frac{\phi_i(\mathbf{r}) \phi_a^*(\mathbf{r}) \phi_i^*(\mathbf{r}') \phi_a(\mathbf{r}')}{\omega + (\epsilon_a - \epsilon_i)} \right]. \quad (2.37)$$

An important property of the exact linear density response is that it diverges (has poles) at the true electronic excitation energies of the unperturbed system. However, note that the true excitation energies are generally not identical to the difference between KS orbital energies $\omega = \epsilon_a - \epsilon_i$ for which $\chi_{\text{KS}}(\omega, \mathbf{r}, \mathbf{r}')$ has poles.⁹

To obtain the true excitation energies of the interacting system, the exact linear response can be parametrized,

$$\rho^{(1)}(\mathbf{r}, \omega) = \sum_{i,a} [X_{ia}(\omega) \phi_a^*(\mathbf{r}) \phi_i(\mathbf{r}) + Y_{ia}(\omega) \phi_a(\mathbf{r}) \phi_i^*(\mathbf{r})], \quad (2.38)$$

with the ground-state KS orbitals. Using this parametrization together with the expression for $\chi_{\text{KS}}(\omega, \mathbf{r}, \mathbf{r}')$ above (Eq. 2.37), the generalized eigenvalue problem

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^* & \mathbf{A}^* \end{pmatrix} \begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix} = \omega \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix} \quad (2.39)$$

is obtained, where the eigenvalue $\omega = \omega_{I0} = (E_I - E_0)/\hbar$ corresponds to the true vertical excitation energy from the ground state to excited state I .⁸ The matrix elements of $\mathbf{A}$ and $\mathbf{B}$, in turn, are defined as

$$A_{ia,jb} = \delta_{ab}\delta_{ij}(\epsilon_a - \epsilon_i) + K_{ia,jb} \quad (2.40)$$

and

$$B_{ia,jb} = K_{ia,bj}, \quad (2.41)$$

where

$$K_{kl,mn} = \int \phi_k^*(\mathbf{r})\phi_l(\mathbf{r}) \left(\frac{1}{|\mathbf{r} - \mathbf{r}'|} + f_{xc}(\mathbf{r}, \mathbf{r}') \right) \phi_n^*(\mathbf{r}')\phi_m(\mathbf{r}')d\mathbf{r}'d\mathbf{r}. \quad (2.42)$$

The exact structures of $\mathbf{A}$ and $\mathbf{B}$ depend on the specific exchange-correlation functional used.⁸

For real orbitals (typically used in practical calculations), the matrix $(\mathbf{A} - \mathbf{B})$ becomes positive definite and the dimensionality of the eigenvalue problem can be reduced by reformulating Eq. 2.39 as

$$(\mathbf{A} - \mathbf{B})^{1/2} (\mathbf{A} + \mathbf{B}) (\mathbf{A} - \mathbf{B})^{1/2} (\mathbf{X} + \mathbf{Y})' = \omega^2 (\mathbf{X} + \mathbf{Y})', \quad (2.43)$$

where $(\mathbf{X} + \mathbf{Y})' = (\mathbf{A} - \mathbf{B})^{1/2} (\mathbf{X} + \mathbf{Y})$.¹¹ This pseudo-eigenvalue equation, proposed by Casida,²² is one of the most popular approaches to calculate excitation energies within TD-DFT.

Eq. 2.43 (or generally Eq. 2.39) provides a route for obtaining all excitation energies in one single calculation without explicitly considering excited-state wave functions. The transition dipole moment between the ground state and excited state I , $\langle\psi_0|\hat{\mu}|\psi_I\rangle$, which can be used to calculate the oscillator strengths, can also be obtained from the ground-state wave function.¹⁰ Furthermore, it is also possible to derive analytical gradients without explicitly considering the excited-state wave functions.^{23–27} The development of analytical gradients in TD-DFT was an important step in the field of computational photochemistry because it enabled systematic exploration of excited-state PESs of large molecules. This also made it possible to obtain photophysical and photochemical properties of such systems beyond vertical excitation energies.

The accuracy of TD-DFT excitation energies depends on both occupied and virtual KS orbitals and energies. With many exchange-correlation functionals, low-lying virtual orbitals are more accurately described than the corresponding high-lying ones. Therefore, TD-DFT normally performs best for low-lying valence excited states.

Today, TD-DFT has become the primary tool for calculating excited states of large molecules. However, there are some well-known deficiencies associated with TD-DFT such as the description of states with charge-transfer or double-excitation character. Although charge-transfer states can normally be described with range-separated hybrid functionals such as LC-BLYP,²⁸ CAM-B3LYP,²⁹ and

ω B97X-D,³⁰ it is also of interest to consider some alternative methods available for studying excited states of large molecules.

2.3 Other Methods for Modeling Excited States

2.3.1 Configuration Interaction Methods

The simplest *ab initio* method for treating excited states is the configuration interaction singles (CIS) approach.⁶ Starting from the optimized HF wave function, the CIS wave function ψ_{CIS} is constructed as a linear combination of all singly excited Slater determinants ψ_i^a that can be obtained by replacing one occupied orbital i with one virtual orbital a (Fig. 2.1). Explicitly

$$\psi_{\text{CIS}} = \sum_{ia} c_i^a \psi_i^a, \quad (2.44)$$

where the expansion coefficients $\{c_i^a\}$ are obtained by variationally minimizing the energy for a set of frozen HF orbitals. Because of Brillouin's theorem ($\langle \psi_{\text{HF}} | \hat{H} | \psi_i^a \rangle = 0$), the CIS wave function is orthogonal to the HF wave function.¹⁵

Figure 2.1. Examples of excited Slater determinants generated from an HF wave function.

CIS typically overestimates excitation energies by more than 1 eV, which means that this method is usually inferior to TD-DFT.^{31,32} The large errors in the CIS energies arise from the neglect of correlation effects together with the fact that the HF molecular orbitals often give a poor representation of “true” molecular orbitals. While CIS for a long time was the standard *ab initio* method for treating excited states of large molecules, some configuration interaction (CI)-based approaches that include correlation effects in the excited states are today available

for modeling large systems such as the CIS with perturbation-theory treatment of doubles corrections (CIS(D)) approach.³³

2.3.2 Coupled Cluster Methods

Another family of methods available for calculating excited states are those based on coupled cluster (CC) theory. These methods can be classified into different groups like the symmetry-adapted cluster configuration interaction (SAC-CI),³⁴ the equation-of-motion CC (EOM-CC)³⁵ and the hierarchy of linear response CCx (CCS, CC2, CC3 ...) methods.³⁶ Although the family of CC methods can systematically be improved to yield very accurate results, only those that include effects from single and double excitations such as SAC-CI and CC2 can be used for calculations on large molecules. While SAC-CI is often the most sophisticated CC method available for such systems, CC2 is among the cheapest methods able to describe electron correlation in excited states.

2.3.3 Multiconfigurational Methods

The final family of *ab initio* methods that will be presented here are the multiconfigurational self-consistent field (MCSCF)-based methods. In general, a MCSCF wave function is constructed as a linear combination of excited Slater determinants (Fig. 2.1), and is obtained by simultaneously optimizing orbitals and expansion coefficients. Since it is usually impossible to do so for all possible Slater determinants, only a subset of them is selected based on some method-dependent criteria. In the complete active space self-consistent field (CASSCF) method, for example, the linear combination of Slater determinants only includes those that arise from a user-specified selection of so-called active orbitals and electrons. This selection has to be governed by some *a priori* knowledge of the character of the state in question, which is a drawback and makes CASSCF a difficult method to use.

While CASSCF captures non-dynamical correlation, it gives a poor description of dynamical correlation. To properly account for both dynamical and non-dynamical correlation, the CASSCF wave function can be used as a starting point (reference function) for carrying out complete active space second-order perturbation theory (CASPT2)¹⁶ calculations. CASPT2 is the most generally applicable and accurate method for the calculation of excited states of small molecular systems. However, although it is regarded as the gold standard for such systems, it is challenging to obtain the same accuracy for large molecules for which large basis sets and large active spaces can not be used. Therefore, if non-dynamical correlation effects are of minor importance, SAC-CI often provides a much better alternative.

2.4 Relativistic Effects

One of the most common approximations in quantum chemistry is that relativistic effects are neglected. Although this is a reasonable approximation in many photochemical applications, for one of the systems in the present thesis (*i.e.*, lumiflavin)

relativistic effects actually play a critical role for the observed photochemistry. In this section, a brief introduction to the field of relativistic quantum chemistry is given. For more complete accounts of relativistic quantum chemistry, the reader is referred to textbooks and reviews in the field such as Refs. 37–41.

2.4.1 The Dirac Equation

In a fully relativistic quantum mechanical description of molecular systems within the BO approximation, stationary electronic states are represented by wave functions, which are eigenfunctions of the time-independent Dirac equation. This equation can be written as

$$\hat{H}\psi = E\psi, \quad (2.45)$$

which is exactly the same general form as the non-relativistic counterpart the Schrödinger equation. However, while a wave function ψ is a scalar function in the non-relativistic framework, it is a four-component (4C) vector in the fully relativistic framework. In the standard representation this wave function is given by

$$\psi = \begin{pmatrix} \psi^{\alpha L} \\ \psi^{\beta L} \\ \psi^{\alpha S} \\ \psi^{\beta S} \end{pmatrix}, \quad (2.46)$$

where α and β are spin labels, and L and S represent the so-called large and small components. The reason why they are called so is that for electronic solutions of the Dirac equation the small components vanish in the non-relativistic limit, whereas the large components converge toward the corresponding non-relativistic wave functions. The Hamiltonian used in Eq. 2.45 can be written in the general form as

$$\hat{H} = \sum_i \hat{h}_D + \frac{1}{2} \sum_{i \neq j} \hat{g}(i, j) + \hat{V}_{\text{en}}, \quad (2.47)$$

where

$$\hat{h}_D = \beta' c^2 + c(\boldsymbol{\alpha} \cdot \mathbf{p}) \quad (2.48)$$

is the one-electron Dirac operator and $\hat{g}(i, j)$ is the two-electron operators accounting for the interaction between electrons i and j . In Eq. 2.48, c is the speed of light, $\mathbf{p}$ is the linear electron momentum, and β' and $\boldsymbol{\alpha} = (\alpha_x, \alpha_y, \alpha_z)$ are the 4C matrices

$$\beta' = \begin{pmatrix} 0 & 0 \\ 0 & -2I_2 \end{pmatrix} \quad \text{and} \quad \alpha_i = \begin{pmatrix} 0 & \sigma_i \\ \sigma_i & 0 \end{pmatrix}, \text{ respectively,} \quad (2.49)$$

constructed from the two-dimensional identity (I_2) and Pauli spin matrices

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad \text{and} \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.^{37-41} \quad (2.50)$$

The relativistic effects to the two-electron interaction have two different origins. First, the electromagnetic interactions are mediated by photons, which travel with a finite velocity (*i.e.*, the speed of light). This means that the interaction is retarded and that an electron will not feel the repulsion from another electron instantaneously such as in the non-relativistic framework. Second, electrons are moving charges and, accordingly, interact with the magnetic field generated by the current and by the spins of the other electrons. In contrast to the one-electron part of the Dirac Hamiltonian, no analytic expression is available for $\hat{g}(i, j)$. However, this operator can be expanded in a perturbative series in c^{-2} as

$$\hat{g}(i, j) = \frac{I_4 \cdot I_4}{r_{ij}} - \frac{c\boldsymbol{\alpha}_i \cdot c\boldsymbol{\alpha}_j}{c^2 r_{ij}} - \frac{(c\boldsymbol{\alpha}_i \cdot \nabla_i)(c\boldsymbol{\alpha}_j \cdot \nabla_j)r_{ij}}{2c^2} + \mathcal{O}(c^{-4}), \quad (2.51)$$

where the first term is the instantaneous Coulomb operator, which together with the one-electron Dirac operator form the so-called Dirac-Coulomb Hamiltonian.³⁷ This is one of most widely used 4C relativistic many-electron Hamiltonians in computational photochemistry. As an example, a relativistic extension (4C-TD-DFT) of the non-relativistic TD-DFT approach has been derived starting from the Dirac-Coulomb Hamiltonian and implemented in relativistic quantum chemical programs.^{42–47} In these implementations, the working equations for the calculation of excitation energies have been obtained in a similar fashion (using response theory and the adiabatic approximation) as the corresponding non-relativistic TD-DFT equations outlined in Section 2.2.^{44–46}

Although the Dirac-Coulomb Hamiltonian is sufficiently accurate for many chemical systems where relativistic effects come into play, some cases requires the inclusion of the second and third terms in Eq. 2.51, which are the so-called Gaunt and gauge operators. These two operators are here written in terms of the relativistic velocity operators $c\boldsymbol{\alpha}_i$ and $c\boldsymbol{\alpha}_j$ for particle i and j , respectively. The sum of the Gaunt and gauge operators gives the so-called Breit operator, which approximately account for the (small) effects from the magnetic and retarded electron-electron interactions.³⁷

2.4.2 Approximate Relativistic Approaches

Despite the fact that much progress have been made in reducing the computational cost in relativistic calculations based on the 4C Dirac equation described in the previous section,^{48,49} such calculations are still much more demanding than the corresponding non-relativistic calculations. For applications to medium- and large-scale molecular systems, a number of alternative approaches have therefore been developed that account for relativistic effects in a more approximate way.

The first approximation one can make is to transform the fully relativistic 4C Dirac Hamiltonian into a two-component Hamiltonian. By keeping only the most important terms, it is possible to effectively reduce the complexity of the computational problem with relatively little loss in accuracy. Examples of two-component Hamiltonians are the zeroth-order regular approximation and Douglas-Kroll-Hess Hamiltonians.^{39,50,51}

Further approximations can often be made for the modeling of organic systems, where the key relativistic correction to the non-relativistic energy is usually the spin-orbit coupling. An alternative route for treating relativistic effects have therefore evolved, which starts from the non-relativistic Schrödinger Hamiltonian and add a spin-orbit coupling operator as a perturbation. Among all spin-orbit operators reported in the literature, one may highlight the full one- and two-electron Breit-Pauli and Douglas-Kroll transformed spin-orbit coupling operators, as well as the effective one-electron mean-field Hamiltonian.⁵²

For the calculation of valence properties like covalent bonds and vibrational frequencies, further computational savings can be achieved by replacing the core orbitals by a relativistic effective core potential. This is the most economic way to include relativistic effects in the calculations. However, it is important make sure that the quality of the results is sufficiently accurate for the problem to be solved.⁵³

2.5 Environmental Effects

So far, the present chapter has focused on the description of electronic structure, and particularly excited states, using quantum chemical methods, without and with the inclusion of relativistic effects. Using such methods, it is possible to give a detailed description of the electronic structure of the system at hand. However, such methods are often restricted to molecules consisting of up to a few hundreds of atoms. This means that unless new approximations are introduced, there are many molecular systems of interest such as proteins that are impossible to study with contemporary computers. In order to afford a detailed description of electronic processes in large molecular systems, one common approach is to use multiscale models, in which the full system is divided into a number of subsystems treated at different levels of theory.

In this section, two multiscale models will be introduced. The first part is concerned with one particular family of continuum solvation models called polarizable continuum models, which are typically used to describe solvent effects. The second part is concerned with the hybrid quantum mechanics/molecular mechanics approach, which is widely used for modeling biological systems such as proteins.

2.5.1 Polarizable Continuum Models

In a polarizable continuum model (PCM), a solvated molecule is described at a quantum mechanical (QM) level of theory inside a cavity surrounded by a polarizable dielectric continuum (characterized by a dielectric constant ϵ) that represents the solvent (Fig. 2.2). In such a model, the solute induces a charge distribution on the cavity surface, which in turn interacts with the solute and polarizes its wave function. A PCM description of the solvent enables calculation of solvated molecules at a low computational cost but only captures bulk electrostatic effects. If specific solute-solvent interactions are important, the solute model can be expanded to a supermolecule by adding explicit solvent molecules.⁴⁰

Figure 2.2. Schematic illustration of the PCM model.

In practical calculations, the solute-solvent boundary surface is subdivided into small elements (tesserae). Each tessera is characterized by an apparent charge q_i and a position $\mathbf{s}_i$. Based on these variables, the electronic perturbation operator can be expressed as

$$\hat{V}_\sigma = \sum_{\text{tesserae}} \frac{q_i}{|\mathbf{s}_i - \mathbf{r}|}, \quad (2.52)$$

where $\mathbf{r}$ represents the electronic coordinates.⁴⁰

In the PCM approach, one can define (omitting thermal motions) an absolute free energy in solution G as

$$G = E[\rho^0] + \Delta G_{\text{solv}}, \quad (2.53)$$

where $E[\rho^0]$ is the energy of the solute in the gas phase with electron density ρ^0 and

$$\Delta G_{\text{solv}} = G_{\text{cav}} + G_{\text{disp-rep}} + G_{\text{el}} - E[\rho^0] \quad (2.54)$$

is the free energy of solvation. In this expression, G_{cav} accounts for the work to create the cavity in the dielectric continuum and $G_{\text{disp-rep}}$ is the dispersion-repulsion term that accounts for the exchange interaction and Pauli repulsion between the solute and solvent electrons. These two terms represents the non-electrostatic part of the free energy.⁴⁰

The electrostatic free energy term G_{el} can be obtained from

$$G_{\text{el}} = E[\rho] + G_{\text{int}}, \quad (2.55)$$

where $E[\rho]$ is the energy in the gas phase with the perturbed density $\rho \neq \rho^0$ and

$$G_{\text{int}} = \frac{1}{2} \sum_i q_i V_i \quad (2.56)$$

is the electrostatic solute-solvent interaction free energy. Here, V_i is the total potential generated by the solute (both electrons and nuclei) in tessera i . The three important variables ρ , q_i and V_i can be determined by variationally minimizing

$$G_{\text{el}} = \langle \psi | \hat{H}^0 + \frac{1}{2} \hat{V}_\sigma | \psi \rangle \quad (2.57)$$

subject to some model specific boundary conditions, where $\hat{H}^0$ is the gas-phase Hamiltonian. The free energy can thereafter be obtained.⁴⁰

Non-Equilibrium Solvation in TD-DFT

The dynamic nature of the solvent plays an important role in many photophysical and photochemical processes such as photon absorption and emission. During such processes, solvent electronic (fast) degrees of freedom have time to respond to the change in solute electron density, whereas solvent nuclear (slow) degrees of freedom are delayed or even frozen. In order to simulate such a non-equilibrium situation, each apparent tessera charge q_i is divided into a fast

$$q_{i,\text{f}} = \frac{\epsilon_{\text{f}} - 1}{\epsilon - 1} q_i \quad (2.58)$$

and a slow component

$$q_{i,\text{s}} = \frac{\epsilon - \epsilon_{\text{f}}}{\epsilon - 1} q_i, \quad (2.59)$$

where $\epsilon_{\text{f}} \neq \epsilon$ is the dielectric constant related to the fast degrees of freedom. In TD-DFT calculations of excited states of solvated molecules, non-equilibrium solvation can be modeled within either the linear-response (LR) or the state-specific (SS) formalism.⁴⁰

Within the LR formalism, the time-dependent KS equations are modified by adding a perturbation $v^{\text{PCM}}(\epsilon)$. The perturbed KS equations then take the form

$$[\hat{H}_{\text{KS}}^0 + v^{\text{PCM}}(\epsilon)]\phi(\mathbf{r}, t) = i \frac{\partial}{\partial t} \phi(\mathbf{r}, t), \quad (2.60)$$

where $\hat{H}_{\text{KS}}^0$ is the time-dependent KS operator in the gas phase. While the v^{PCM} potential depends on ϵ if both fast and slow degrees of freedom have time to relax (equilibrium solvation), it depends on ϵ_{f} only if only the fast degrees of freedom has time to relax (non-equilibrium solvation).⁵⁴

In the SS implementation of PCM/TD-DFT, the electrostatic interaction part of the free energy of the excited state explicitly depends on the ground-state density and is given by the expression

$$\begin{aligned} G_{\text{int,SS,neq}}^{(2)} &= \frac{1}{2} \sum_i q_{i,\text{f}}^{(2)} V_{i,\rho}^{(2)} + \left(\sum_i q_{i,\text{s}}^{(1)} V_{i,\rho}^{(2)} - \frac{1}{2} \sum_i q_{i,\text{s}}^{(1)} V_{i,\rho}^{(1)} \right) \\ &\quad + \frac{1}{2} \left(\sum_i q_{i,\text{s}}^{(1)} V_{i,\text{f}}^{(2)} - \sum_i q_{i,\text{s}}^{(1)} V_{i,\text{s}}^{(1)} \right), \end{aligned} \quad (2.61)$$

where (2) and (1) represent the excited and ground states, respectively. In Eq. 2.61, some of the potentials and tessera charges depend on the excited state electron density and, accordingly, these have to be determined using an iterative procedure. Furthermore, V_f and V_s correspond to the potentials relative to the fast and slow degrees of freedom, respectively, and $V_\rho^{(n)}$ is the potential generated by the density of state (n). Due to the iterative procedure, the SS method is often more computationally demanding than the LR approach. However, at least for polar solvents, the SS approach usually give more accurate results.⁵⁵

2.5.2 QM/MM Methods

The hybrid quantum mechanics/molecular mechanics (QM/MM) approach¹² has become a invaluable tool in modeling complex biomolecular systems such as biological photoreceptors, enzymes and DNA. In the QM/MM approach, the system (**S**) of interest is divided into an inner (**I**) subsystem treated at a QM level of theory and an outer (**O**) subsystem modeled with a molecular mechanics (MM) force field. This is illustrated in Fig. 2.3.

Figure 2.3. Schematic partitioning of the entire system (**S**) into inner (**I**) and outer (**O**) subsystems. The blue ring represents the boundary region.

One of the most difficult tasks in setting up a QM/MM calculation is to select **I** and **O**. The general principle is that the chemically active part of the system (such as the active site of a protein) should be included in **I** (the QM part), whereas the remaining part of the system can form **O** (the MM part). At the boundary between **O** and **I**, the atoms interact with other atoms included in both subsystems. Therefore, it is important to achieve a well-balanced description of all types of interactions in this region.

In many cases, it is unavoidable that the boundary between **I** and **O** cuts across covalent bonds. To minimize possible artifacts for the description of the

QM region, the dangling bonds of the QM atoms covalently attached to the MM region have to be capped. This can, for example, be done by adding link atoms (typically hydrogen atoms) or by using frozen orbitals.⁵⁶

Having partitioned $\mathbf{S}$ into $\mathbf{I}$ and $\mathbf{O}$, the QM/MM energy ($E_{\text{QM/MM}}$) of $\mathbf{S}$ can be obtained from either an additive

$$E_{\text{QM/MM}}^{\text{add}}(\mathbf{S}) = E_{\text{MM}}(\mathbf{O}) + E_{\text{QM}}(\mathbf{I}+\mathbf{L}) + E_{\text{QM-MM}}(\mathbf{I},\mathbf{O}) \quad (2.62)$$

or a subtractive

$$E_{\text{QM/MM}}^{\text{sub}}(\mathbf{S}) = E_{\text{MM}}(\mathbf{S}) + E_{\text{QM}}(\mathbf{I}+\mathbf{L}) - E_{\text{MM}}(\mathbf{I}+\mathbf{L}). \quad (2.63)$$

scheme. Here, $\mathbf{L}$ and $\mathbf{I}+\mathbf{L}$ represent link atoms and the capped QM system, respectively.⁵⁶

The ONIOM Approach

The ONIOM (Our N-layered Integrated molecular Orbital and molecular Mechanics) approach⁵⁷ is a popular method for carrying out QM/MM calculations and is based on the subtractive scheme. Although ONIOM can be generalized to treat an arbitrary number of QM and/or MM layers, only the two-layer ONIOM(QM:MM) scheme will be discussed here.⁵⁸

In the original formulation of ONIOM(QM:MM), the $E_{\text{QM/MM}}^{\text{sub}}(\mathbf{S})$ energy is obtained from the $E_{\text{QM}}(\mathbf{I}+\mathbf{L})$ energy computed with a QM method, and the $E_{\text{MM}}(\mathbf{O})$ and $E_{\text{MM}}(\mathbf{I}+\mathbf{L})$ energies calculated with an MM force field. Adopting the notation by Cornell et. al,⁵⁹ a typical force field energy can be expressed as

$$E_{\text{MM}} = E_r + E_\theta + E_\phi + E_{\text{vdW}} + E_{\text{el}}, \quad (2.64)$$

where the bond stretching term is given by

$$E_r = \sum_{\text{bonds}} K_r (r - r_{\text{eq}})^2, \quad (2.65)$$

the bond bending term by

$$E_\theta = \sum_{\text{angles}} K_\theta (\theta - \theta_{\text{eq}})^2, \quad (2.66)$$

and the dihedral angle term by

$$E_\phi = \sum_{\text{dihedrals}} \frac{V_n}{2} (1 + \cos(n\phi - \gamma)). \quad (2.67)$$

The final two terms

$$E_{\text{vdW}} = \sum_{i < j} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} \right] \quad (2.68)$$

and

$$E_{\text{el}} = \sum_{i < j} \frac{q_i q_j}{\epsilon R_{ij}} \quad (2.69)$$

account for the van der Waals and the electrostatic interactions, respectively. In the original ONIOM approach, the electrostatic coupling term between the **I** and **O** regions is treated at the MM level of theory, which means that the QM electron density and nuclear charges has been replaced by fixed atomic charges (so-called mechanical embedding). Normally, this is a very crude approximation for photochemical reactions.

A more accurate way of treating the electrostatic QM-MM coupling term is provided by the electrostatic embedding (EE) scheme.⁶⁰ In this embedding scheme, point charges (normally taken from the MM force field) are incorporated into the QM Hamiltonian, which is thus augmented by a term

$$\hat{H}^{\text{EE}} = - \sum_{i=1}^N \sum_{J=1}^{M'} \frac{s_J q_J}{|\mathbf{R}_J - \mathbf{r}_i|} + \sum_{I=1}^M \sum_{J=1}^{M'} \frac{Z_I s_J q_J}{|\mathbf{R}_J - \mathbf{R}_I|}. \quad (2.70)$$

In this expression, the scaling factor s_J prevents severe overpolarization of the QM wave function, q_J is the MM point charge located at $\mathbf{R}_J$, Z_I is the nuclear charge of QM nucleus I , and $\mathbf{R}_I$ and $\mathbf{r}_i$ represent the nuclear and electronic coordinates, respectively. The indices i , I and J run over the N QM electrons, the M QM nuclei, and the M' MM atoms. Since the additional term in the QM Hamiltonian captures the electrostatic interactions between the two subsystems, one has to remove the corresponding classical Coulomb interactions from the total energy. This is done by adding a classical Coulomb term to the $E_{\text{MM}}(\mathbf{I}+\mathbf{L})$ energy, and the modified Coulomb term then becomes

$$E_{\text{MM}}^{\text{EE}}(\mathbf{I}+\mathbf{L}) = E_{\text{MM}}(\mathbf{I}+\mathbf{L}) + \sum_{I=1}^M \sum_{J=1}^{M'} \frac{q_I s_J q_J}{|\mathbf{R}_J - \mathbf{R}_I|}. \quad (2.71)$$

Here, the scaling factor s_J is identical to that in Eq. 2.70. Including electrostatic embedding, the QM/MM energy finally reads

$$E_{\text{QM/MM}}^{\text{sub,EE}}(\mathbf{S}) = E_{\text{MM}}(\mathbf{S}) + E_{\text{QM}}^{\text{EE}}(\mathbf{I}+\mathbf{L}) - E_{\text{MM}}^{\text{EE}}(\mathbf{I}+\mathbf{L}). \quad (2.72)$$

Although electrostatic embedding is a substantial improvement compared with the classical mechanical embedding scheme, the general issue concerning the compatibility between the MM charges and the QM electron density still remains. Inherent in the electrostatic embedding scheme is the assumption that the point charges give a good representation of the real charge distribution of the MM system. However, the MM point charges are typically designed together with all other force field parameters to give a balanced description of all types of interactions. Another potential source of concern is how the QM method responds to point charges. For example, it has been shown for rhodopsins that DFT methods are not able to fully account for the electrostatic effects.⁶¹ Despite these concerns, the QM/MM methodology has today become the state-of-the-art computational technique for modeling complex biomolecular systems.

2.6 Concluding Remarks

In this chapter, a number of methods have been presented that can be used to solve the non-relativistic electronic Schrödinger equation (or the relativistic electronic Dirac equation). Employing one of them to calculate the electronic energy of the system for various sets of nuclear coordinates, a stationary description of the most representative molecular geometries along a PES can be obtained. While such an analysis is sufficient in many photochemical applications, others requires that also the dynamics of the nuclei are taken into account. Although this thesis is concerned with solving the (static) electronic problem, it is nevertheless worth mentioning that there are of course methods available for studying the nuclear problem such as trajectory surface hopping methods,⁶² where the nuclei are propagated on a surface according to Newton's equations of motions and the system is able to switch between different electronic states.

CHAPTER 3

Photobiological Keto-Enol Reactions

Many biological systems rely on a careful control of the relative concentrations of keto and enol forms (usually referred to as keto and enol tautomers) of key molecules. DNA, for example, is exclusively built from nucleobases in their keto tautomeric forms, and the conversion of a single base into its enol form is a common source of mutations.⁶³ The underlying control is in many cases executed by enzymes that are able to rapidly adjust the keto and enol concentrations by catalyzing the interconversion of the two forms.⁶⁴

Following an introduction to keto-enol tautomerism, this chapter presents my research on two photobiologically relevant organic chromophores for which keto-enol reactions play important roles. The two chromophores are astaxanthin and oxyluciferin. Astaxanthin is responsible for the color of lobster carapace and oxyluciferin is responsible for the light emission of fireflies.

3.1 Keto-Enol Tautomerism

Figure 3.1. Keto-enol tautomerism.

The conversion of the parent keto form of a carbonyl containing organic molecule

(a ketone or aldehyde) into its enol form involves the shuttling of an hydrogen atom (or a proton) from the α -carbon to the carbonyl oxygen, and is accompanied by the saturation of the carbonyl bond by an adjacent single bond (Fig. 3.1). This conversion can be catalyzed by either an acid or a base. The acid-catalyzed keto-enol tautomerism (Fig. 3.2) is initiated by protonation of the carbonyl oxygen atom, which activates the α -hydrogens. Thereafter, the resulting intermediate loses one of the activated hydrogens to yield the enol form. In most cases, the second step is slower than the first.^{65–67}

Figure 3.2. Acid-catalyzed keto-enol tautomerism.

The base-catalyzed keto-enol tautomerism (Fig. 3.3), in turn, starts with deprotonation of the acidic α -carbon to give an enolate intermediate, followed by protonation of the carbonyl oxygen to produce the enol form.^{65–67} Since the enolate form is resonance-stabilized, a carbonyl compound is usually a much stronger carbon acid than is an alkane. As an illustrative example, the pK_a value for acetone is ~ 40 units smaller than the pK_a value for ethane (Fig. 3.4).⁶⁷ Analogous to the acid-catalyzed reaction, it is the deprotonation of the α -carbon that determines the rate for the overall base-catalyzed reaction.^{65–67}

Figure 3.3. Base-catalyzed keto-enol tautomerism.

There are many factors that affect the equilibrium between keto and enol tautomers. Although a carbonyl group is intrinsically more stable than an enol group, the possibility of additional stabilization effects from other functional groups may shift the equilibrium in favor of the enol tautomer.^{68–70} Phenol, for example, is

Figure 3.4. Chemical structures of acetone and ethane.

much more stable than the corresponding keto tautomer (2,4-cyclohexadienone) because of aromaticity (Fig 3.5).^{65,66} Vinyl alcohol, which is stabilized by both resonance and intramolecular hydrogen bonding, is another stable enol (Fig 3.5). In fact, while the parent keto acetylacetone form is slightly more stable than vinyl alcohol in aqueous solution, the situation is actually reversed in nonpolar solvents like tetrachloromethane that promotes intramolecular interactions.⁷¹ Other factors that may shift the equilibrium toward the enol form (or the keto form) are electron withdrawing/donating groups, steric effects, and temperature.^{65–67}

Figure 3.5. Chemical structures of 2,4-cyclohexadienone, phenol, acetylactone and vinyl alcohol.

Since the tautomeric equilibria in excited states can be very different from those in the ground state, it is in many cases possible to induce an isomerization between two tautomeric forms by light.^{72–79} Already in 1961, Yang and Rivas showed that the keto form of *ortho*-methylbenzophenone (Fig. 3.6) can undergo excited-state tautomerism into the corresponding enol form in alcohol solutions. Furthermore, Naumov and co-workers have recently used time-resolved emission spectroscopy to show that photoexcitation can induce enolization of the keto form of 6'-dehydroxyoxyluciferin (Fig. 3.6).⁷⁹ Several examples of photoketonization have also been reported in the literature.^{75–78} For example, Yankov *et al.* have shown that the keto form of dibenzoylmethane (Fig. 3.6), which is unstable in the ground state, is formed upon photoexcitation.⁷⁶

After this brief introduction to keto-enol tautomerism, we will in the following two sections describe two organic chromophores that can exist in a number of

Figure 3.6. Chemical structures of the keto and enol forms of *ortho*-methylbenzophenone, 6'-dehydroxyoxyluciferin and dibenzoylmethane.

ketonic and enolic forms. The first chromophore is astaxanthin, which will be discussed in the next section, and the second one is oxyluciferin, discussed in the subsequent section.

3.2 Astaxanthin

The deep-blue color of live lobster arises from a protein complex present in the lobster carapace. This complex is called α -crustacyanin and consists of eight β -crustacyanin dimers, where each dimer contains two astaxanthin (3,3'-dihydroxy- β -carotene-4,4'-dione, AXT, Fig. 3.7) chromophores.⁸⁰ Upon cooking of live lobsters, α -crustacyanin denatures to leave the AXT molecules unbound in solution. While AXT has a UV-vis absorption maximum (λ_{max}) around 630 nm in α -crustacyanin (580 – 590 nm in β -crustacyanins), the free chromophore in solution absorbs in the 475 – 500 nm region, giving the well-known orange-red color of cooked lobster.^{81–83} Thus, the protein environment induces a > 0.5 eV bathochromic shift of the absorption by AXT, which is one of the largest protein-induced spectral shifts recorded in nature.

To date, a number of different mechanisms have been put forward to explain the chemical basis for this shift. These mechanisms follow two separate themes. The first is concerned with the physical arrangements of the AXT chromophores inside the protein complexes. Although an atomic-resolution structure of the full

Figure 3.7. Chemical structure of the neutral α -hydroxyketone form of astaxanthin (AXT).

α -crustacyanin complex is not available, Helliwell and co-workers⁸⁴ have reported X-ray crystallographic data of a β -crustacyanin that reveal that the binding of AXT to the apoprotein both changes the conformation of the chromophore from 6-*s-cis* to 6-*s-trans*, and reduces the twist of the β -ionone end-rings relative to the polyene chains about the C6-C7 (and C6'-C7') bonds (see atom numbering in Fig. 3.7). Since co-planarization of the end-rings with the polyene chain increases the effective conjugation of the chromophore, it follows that the absorption should be bathochromically shifted in the protein. The β -crustacyanin crystal structure also shows that the two AXT chromophores are paired with their polyene chains lying quite close to each other (~ 7 Å), which would facilitate an exciton-coupling mechanism. However, both computational and experimental studies have shown that neither conformational changes nor exciton coupling can account for more than one third of the observed shift.⁸⁵⁻⁸⁷

The second theme, which was suggested already in 1938 by Kuhn and Sørensen,⁸⁸ is concerned with the keto-enol and acid-base chemistry of AXT. Specifically, although the neutral α -hydroxyketone form of AXT (**1** in Fig. 3.8) is predominant in solution, these authors suggested that the shift arises as a consequence of the bias of the apoprotein to rather bind AXT in its anionic enolate form (**1**⁻). While such a mechanism has remained largely neglected, **Paper I** presents experimental and computational results showing that this scenario is indeed possible and associated with a large bathochromic shift.

3.2.1 Paper I

In order to evaluate the possibility that the bathochromic shift is due to the presence of AXT in its enolate form **1**⁻ in α -crustacyanin, the absorption maximum of the parent keto form **1** was first calculated with TD-DFT employing three different functionals (M06-HF, LC- ω PBE and LC-BLYP) and two different basis sets (6-31G(d,p) and aug-cc-pVDZ) in water and methanol solutions. These results were then compared with the corresponding absorption maxima obtained for the mono-enol (**1ed**), di-enol (**1ded**), enolate (**1**⁻) forms shown in Fig. 3.8.

Figure 3.8. Chemical structure of the neutral α -hydroxyketone form of AXT (**1**) and acid-base (HA and B) induced tautomerism of this form producing the enolate (**1⁻**), mono-enol (**1ed**) and di-enol (**1ded**) forms. Reprinted from **Paper I** with permission from the PCCP Owner Societies.

Table 3.1. Calculated absorption maxima ($\lambda_{\max}$, in nm), excitation energies (ΔE , in eV) and oscillator strengths (f) of different forms of AXT.^a Reprinted from **Paper I** with permission from the PCCP Owner Societies.

Method/basis set ^b		1	1ed	1ded	1⁻
Water solvent					
M06-HF/I	$\lambda_{\max}$	457	463	459	660
	ΔE	2.71	2.68	2.70	1.88
	f	4.39	4.48	4.56	4.04
LC- ω PBE/I	$\lambda_{\max}$	473	482	477	658
	ΔE	2.62	2.57	2.60	1.88
	f	4.53	4.64	4.73	4.43
LC-BLYP/I	$\lambda_{\max}$	463	470	465	643
	ΔE	2.68	2.64	2.66	1.93
	f	4.59	4.70	4.79	4.45
M06-HF/II ^c	$\lambda_{\max}$	470	476	471	666
	ΔE	2.64	2.60	2.63	1.86
	f	4.31	4.37	4.44	4.17
LC- ω PBE/II ^c	$\lambda_{\max}$	482	490	484	665
	ΔE	2.57	2.53	2.56	1.86
	f	4.47	4.55	4.64	4.45
LC-BLYP/II ^c	$\lambda_{\max}$	474	481	475	651
	ΔE	2.62	2.58	2.61	1.90
	f	4.52	4.61	4.69	4.50
Methanol solvent					
M06-HF/I	$\lambda_{\max}$	456	463	459	668
	ΔE	2.72	2.68	2.70	1.86
	f	4.39	4.48	4.56	4.05
LC- ω PBE/I	$\lambda_{\max}$	473	481	477	666
	ΔE	2.62	2.58	2.60	1.86
	f	4.52	4.64	4.73	4.45
LC-BLYP/I	$\lambda_{\max}$	463	470	465	651
	ΔE	2.68	2.64	2.67	1.90
	f	4.59	4.70	4.79	4.46
Mixture					
Experiment ^d	$\lambda_{\max}$	476	—	—	—
	ΔE	2.61	—	—	—
	$\epsilon_{\max}^e$	10000	—	—	—

^aAll calculations based on geometries optimized using B3LYP/6-31G(d,p) in combination with a CPCM description of the water or methanol solvent. ^bBasis set I = 6-31G(d,p); basis set II = aug-cc-pVDZ. ^cSolvent effects obtained using the cc-pVDZ basis set and added to gas-phase data obtained with the target aug-cc-pVDZ basis set. ^dExperimental value in a water-methanol mixture. ^e $\epsilon_{\max}$ given in units of $\text{dm}^3 \text{mol}^{-1} \text{cm}^{-1}$

Starting with the parent **1** species, we note from Table 3.1 that all three density functionals used for the calculations give very similar results, and that the experimental value (476 nm) obtained in a water-methanol mixture is reproduced

with an accuracy of ~ 20 nm (~ 0.1 eV). Turning to the calculated absorption maxima of the other forms of AXT, it can be seen that neither single ($\mathbf{1} \rightarrow \mathbf{1ed}$) nor double ($\mathbf{1} \rightarrow \mathbf{1ded}$) enolization is likely to significantly shift the absorption of $\mathbf{1}$. Indeed, regardless of the level of theory, it is predicted that such processes induce shifts smaller than 0.05 eV. Notably, however, the calculations predict that the conversion from $\mathbf{1}$ to $\mathbf{1}^-$ is associated with a bathochromic shift in the 0.72 – 0.86 eV range. Interestingly, this shift is larger than that in β -crustacyanin and similar to that in α -crustacyanin. This gives support for the hypothesis that the bathochromic shift in α -crustacyanin is due to the presence of AXT in its enolate form $\mathbf{1}^-$. However, how can this result be validated experimentally?

Figure 3.9. Chemical structures of 3-hydroxy-4-oxo- β -ionone **2** and its ene-diol tautomer **2ed**, enolate anion $\mathbf{2}^-$ and dianion **2-dianion**. Reprinted from **Paper I** with permission from the PCCP Owner Societies.

Since similar aqueous solution experiments are not possible, due to aggregate formation and the low solubility of AXT in water, the behavior of a water-soluble 3-hydroxy-4-oxo- β -ionone model of AXT (**2**) was examined using both experimental and computational methods. Importantly, the acid-base chemistry of this model compound shown in Fig. 3.9, which retains the trimethylated α -hydroxycyclohexenone of AXT but replaces the rest of the molecule with a simpler conjugated ketone at the 6-position of the ring, is expected to mimic that of AXT.

Comparing the calculated absorption maxima of the various forms of the model compound (**2**, **2ed**, $\mathbf{2}^-$ and **2-dianion**, shown in Fig. 3.9) with the corresponding experimental maxima, it was found that the calculations accurately reproduce the experimental values for **2** and **2-dianion**, and are in qualitative agreement with the experiments in that the formation of the enolate $\mathbf{2}^-$ anion from **2** is

associated with a substantial bathochromic shift. These results, together with complementary benchmark calculations using CC2 and CASPT2, suggest that TD-DFT calculations constitute a reliable tool for obtaining a qualitative accurate assessment of how acid-base chemistry influences the absorption maxima of AXT and models thereof.

Having found that the formation of 1^- induces a large bathochromic shift, it is of interest to investigate whether this process is at all possible in crustacyanins. While a detailed analysis of the stability of the various forms of AXT in different environments is beyond the scope of **Paper I**, both calculations and experiments found that the pK_a of **2** is ~ 10 . Since the conjugation is more extensive in 1^- than in 2^- , the former anion should be more stable and present at even lower pH . In this regard, it is worth noticing that much simpler compounds also having enolate ions in which the negative charge is distributed between two oxygen atoms such as 2,3-dihydroxy-2-propenal (triose reductone) and ascorbic acid have pK_a values below 7 (Fig. 3.10).^{89,90} Further support for the hypothesis that the 1^- form is bound to crustacyanins comes from preliminary B3LYP calculations showing that a key histidine residue shifts the equilibrium between 1^- and **1** by at least 8 kJ mol $^{-1}$ toward the former species. Overall, then, the calculations and experiments presented in **Paper I** show that the formation of 1^- from **1** is associated with a large bathochromic shift and is possible in the protein environment.

Figure 3.10. Chemical structures of 2,3-dihydroxy-2-propenal and ascorbic acid, and their conjugate bases.

3.3 Oxyluciferin

Figure 3.11. Chemical structures of different forms of oxyluciferin and the excited-state equilibrium constants for the keto-enol [$pK_E(S_1)$] and acid-base [$pK_a(S_1)$] reactions that connect them. Reprinted from **Paper II** with permission from John Wiley & Sons. Copyright 2014 Wiley Periodicals, Inc.

The characteristic glow of a firefly derives from a bioluminescent reaction that occurs in the catalytic center of the enzyme firefly luciferase. In this reaction, the ground-state heterocyclic acid D-luciferin is converted into the chemiexcited (S_1 , first excited singlet state) oxyluciferin (OxylH₂) light emitter, which subsequently emits light.^{91,92}

Depending on the chemical conditions, the OxylH₂ light emitter can exist in any of the six spectrally overlapping chemical forms shown in Fig. 3.11. There are two neutral, three mono-anionic and one di-anionic forms. The neutral forms are keto-OxylH₂ and enol-OxylH₂, the mono-anionic forms are phenolate-keto-OxylH⁻, phenolate-enol-OxylH⁻ and enolate-OxylH⁻, and the di-anionic form is OxyL²⁻.^{93,94}

To date, there is no consensus regarding which of the different forms is the main contributor to the light emission *in vivo*. While both computational^{95–97} and experimental^{98,99} studies have proposed that the phenolate-keto-OxylH⁻ form is the preferred light emitter, Naumov and co-workers⁷⁹ have recently studied an OxylH₂ analog (HOxylH) and recorded time-resolved emission data indicating that it is rather an enolate species (enolate-OxylH⁻ or OxyL²⁻) that is responsible for the light emission.

In **Paper II**, the intrinsic tendency of the light emitter to prefer one particular form of OxyLH₂ was investigated by calculating the excited-state pK_E and pK_a values in aqueous solution for the reactions (shown in Fig. 3.11) connecting the different forms of OxyLH₂.

3.3.1 Paper II

Figure 3.12. Structure of the phenolate-keto-OxyLH[−]-water complex considered. Reprinted from **Paper II** with permission from John Wiley & Sons. Copyright 2014 Wiley Periodicals, Inc.

Using a thermodynamic Born-Haber (BH) cycle, pK values in ground [$pK^{BH}(S_0)$] and excited states [$pK^{BH}(S_1)$] were derived by calculating aqueous solution Gibbs free energies of reactants (ketones/acids) and products (enols/bases) in the two states, respectively. In order to minimize inevitable computational errors, the $pK(S_1)$ values, henceforth denoted $pK^{BH,corr}(S_1)$, were obtained by adding calculated [$pK^{BH}(S_1) - pK^{BH}(S_0)$] values to experimental ground-state pK values [$pK^{exp}(S_0)$]

$$pK^{BH,corr}(S_1) = pK^{exp}(S_0) + [pK^{BH}(S_1) - (pK^{BH}(S_0))]. \quad (3.1)$$

Experimental $pK^{exp}(S_0)$ values were taken from Ref. 94 or derived from calculated and experimental data. All calculations were carried out at the $\omega B97X-D/6-31+G(d,p)$ level of theory using a hybrid cluster-continuum model including 11 explicit water molecules. The explicit water molecules were placed as shown for the phenolate-keto-OxyLH[−] anion in Fig. 3.12. The computed $pK^{BH,corr}(S_1)$ and experimental $pK^{exp}(S_0)$ are presented in Fig. 3.13.

Starting by comparing keto-OxyLH₂ with enol-OxyLH₂ (reaction I in Fig. 3.13), we observe that while they are of similar stability in the ground state ($pK_E^{exp}(S_0) = -0.39$), the $pK_E^{BH,corr}(S_1)$ value of ~ 5 is a clear indication that keto-OxyLH₂ is much more stable than enol-OxyLH₂ in the excited state. We also observe that the estimated $pK_a^{BH,corr}(S_1)$ value for the equilibrium reaction between keto-OxyLH₂ and its conjugate base phenolate-keto-OxyLH[−] (reaction IV) is ~ 2 . Since the reference conditions implicated in the interpretation of pK_a values in **Paper II** correspond to a buffered aqueous solution at pH 7, it seems likely that phenolate-keto-OxyLH[−] is favored over keto-OxyLH₂ in the excited state. Hence, neither

Figure 3.13. Experimental ground-state and calculated excited-state equilibrium constants for the keto-enol and acid-base reactions of OxyLH₂. Reprinted from **Paper II** with permission from John Wiley & Sons. Copyright 2014 Wiley Periodicals, Inc.

enol-OxyLH₂ nor keto-OxyLH₂ seems to be a significant contributor to the light emission.

Having predicted that the neutral forms are of minor importance for the light emission, we then proceeded by comparing the three mono-anionic forms. As for the phenolate-keto-OxyLH⁻ $\rightleftharpoons$ enolate-OxyLH⁻ equilibrium (reaction II), it can be seen that while the enolate-OxyLH⁻ form dominates in the ground state, the calculations suggest that it is actually phenolate-keto-OxyLH⁻ that is more stable in the excited state ($pK_E^{\text{BH,corr}}(S_1) \sim 6$). For the phenolate-keto-OxyLH⁻ $\rightleftharpoons$ phenolate-enol-OxyLH⁻ equilibrium (reaction III), in turn, the $pK_E^{\text{BH,corr}}(S_1)$ value of ~ 7 indicates that the phenolate-keto-OxyLH⁻ form is also favored over this enolic mono-anion. Hence, out of the three mono-anionic forms, only phenolate-keto-OxyLH⁻ appears to come into play.

The results presented above show that phenolate-keto-OxyLH⁻ is intrinsically more stable than keto-OxyLH₂, enol-OxyLH₂, phenolate-enol-OxyLH⁻ and enolate-OxyLH⁻ in the excited state in aqueous solution, which means that either phenolate-keto-OxyLH⁻ or OxyL²⁻ is the preferred excited-state species. Since the $pK_a^{\text{BH,corr}}(S_1)$ value for the phenolate-keto-OxyLH⁻ $\rightleftharpoons$ OxyL²⁻ equilibrium (reaction VII) is ~ 14 , it seems likely that phenolate-keto-OxyLH⁻ is also more

stable than OxyL^{2-} in the excited state. Thus, the key conclusion afforded by the excited-state equilibrium constants reported in **Paper II** is that phenolate-keto- OxyLH^- is intrinsically the preferred form of OxyLH_2 in the excited state, which suggests that this species is involved in the bioluminescence of fireflies.

CHAPTER 4

Biological Photoreceptors

Sunlight is an important energy source in Nature. For many organisms, sunlight is also an important environmental stimulus for activating physiological and developmental processes. Specialized proteins, called biological photoreceptors, have therefore evolved that detect light and respond to changes in the ambient light environment. In this chapter, a general overview of these proteins will first be given. Thereafter, my research related to the LOV (light-oxygen-voltage)-domain proteins and the phytochromes, which are two classes of photoreceptors, will be presented.

4.1 Overview

Biological photoreceptors are involved in signal transduction and utilize light energy to activate physiological and developmental processes in all kingdoms of life. Each photoreceptor uses a chromophore to detect light of a specific wavelength within the UV to near-infrared region of the electromagnetic spectrum. The chromophore is either a small organic molecule or is formed from amino-acid residues of the protein. As shown in Fig. 4.1 and Table 4.1, UVR8 (UV resistance locus 8) photoreceptors employ a tryptophan to respond to UV-B light; xanthopsins use coumaric acid to sense blue light; cryptochromes as well as BLUF (blue-light sensors using flavin-adenine dinucleotide) and LOV-domain proteins employ a flavin chromophore to absorb blue light; rhodopsins use retinals to detect light in the UV to visible red spectral range; phytochromes bind a linear tetrapyrrole (bilin) chromophore to detect light in the red and near-infrared regions; and cyanobacteriochromes finally contain a bilin chromophore that absorbs light of UV to near-infrared wavelengths.^{100–104}

The biological activity of a photoreceptor depends on the relative concentra-

Figure 4.1. Spectral sensitivity ranges for different classes of biological photoreceptors.

Table 4.1. Chromophores and primary photochemical events for different classes of biological photoreceptors.

class	chromophore	primary photochemical event
UVR8	tryptophan ^a	proton-coupled electron transfer ^a
xanthopsins	coumaric acid ^b	isomerization ^b
cryptochromes	flavin-adenine dinucleotide (FAD) ^c	electron transfer ^c
BLUF	flavin-adenine dinucleotide (FAD) ^c	rearrangement of hydrogen bonds ^c
LOV	flavin mononucleotide (FMN) ^c	FMN-cysteiny adduct formation ^c
rhodopsins	retinal ^b	isomerization ^b
phytochromes	linear tetrapyrrole (bilin) ^d	isomerization ^d
cyanobacteriochrome	linear tetrapyrrole (bilin) ^d	isomerization ^d

^aRef. 103. ^bRef. 100. ^cRef. 102. ^dRef. 101.

tions of dark-adapted inactive and biologically active forms. Upon light absorption by the inactive form, the photoreceptor undergoes a photocycle (Fig. 4.2), which is started by a photochemical reaction of the chromophore (the so-called *primary photochemical event*, Table 4.1) and involves the formation of the biologically active form via a number of protein intermediate states. The active form, which is typically produced within microseconds after photoexcitation, passes on a signal to other regulatory systems that activate biological pathways in the host organism leading to physiological responses.¹⁰⁵ Since the active form is less stable than the inactive form, a thermally driven spontaneous process called dark reversion slowly (milliseconds to many hours) recovers the inactive form, which closes the photocycle.¹⁰⁴ In addition to the dark reversion, photochromic photoreceptors like

phytochromes have also developed mechanisms to rapidly convert back to the inactive form through light absorption by the active form. Usually, the inactive and active forms of photochromic photoreceptors have different absorption maxima, which enables a rapid interconversion between the two forms in suitable light conditions.^{100,101,104}

Figure 4.2. Schematic illustration of the photocycle of a biological photoreceptor. Green font corresponds to pathways relevant for all photoreceptors. Blue font corresponds to pathway relevant only for photochromic photoreceptors. I_k is intermediate protein state k , and $h\nu$ and Δ indicate light and thermally-induced reaction pathways between the two forms, respectively.

Biological photoreceptors have evolved to maximize their quantum yields for the light-induced chromophore reactions that start the conversion of the inactive form into the active form and, correspondingly, to minimize the quantum yields for competing non-productive processes such as fluorescence. However, although the quantum yield for fluorescence in natural photoreceptors is usually small, it is possible to substantially increase it by introducing point mutations or incorporating chromophore analogs that cannot undergo the biologically productive photoreaction. For example, through mutations of key residues, phytochromes have been engineered into fluorescent probes.¹⁰⁶ LOV-domain proteins have also been made strongly fluorescent by mutating a cysteine residue important for the photoconversion.¹⁰⁷ Since many fluorescent proteins derived from biological photoreceptors have properties that are not easily attained by other fluorescent proteins like the green fluorescent protein (GFP) from jellyfish and its derivatives, they constitute suitable templates for the development of novel imaging tools in biology.¹⁰⁸ An alternative application of biological photoreceptors lies in optogenetics, which is a technology whereby different biological processes are controlled, at a molecular level, using genetically encoded light-sensitive proteins.^{109,110}

After this brief introduction to biological photoreceptors, the following sections will focus on the LOV-domain protein and phytochrome photoreceptor families.

4.2 LOV-Domain Proteins

Figure 4.3. Chemical structure of the flavin mononucleotide (FMN) chromophore.

The LOV-domain proteins are a widespread superfamily of blue-light-sensing biological photoreceptors that play key roles in the regulation of physiological processes such as phototropism^{111–113} and chloroplast relocation^{114,115} in plants. While the flavin mononucleotide (FMN) chromophore (Fig. 4.3) is non-covalently bound to the apoprotein in the dark-adapted inactive form of the photoreceptor, the biologically active form contains a covalent bond between the chromophore and a nearby cysteine residue (Fig. 4.4).

As shown in Fig. 4.4 and described in detail elsewhere,^{116,117} the conformational change from the inactive to the active form of the photoreceptor begins with a transition from the electronic ground state to the lowest electronically excited singlet state (S_1). Thereafter, a rapid relaxation from the S_1 state to the lowest electronically excited triplet state (T_1) occurs via an intermediate triplet state (T_n). Then, the covalent bond between the chromophore and the cysteine residue is formed, which induces structural changes in the protein leading to the formation of the biologically active form of the photoreceptor.

In **Paper III**, the transition between the S_1 and T_n states was used as an example to demonstrate the ability of first-principles methods to calculate rate constants for inter-excited state phosphorescence events. While it is well-established that first-principles calculations can be carried out to determine rate constants for intersystem crossing events between excited states, this was the first time that such calculations were performed to obtain rate constants for the competing phosphorescence channels.

Figure 4.4. Photocycle of LOV-domain proteins ($R = \text{CH}_2(\text{CHOH})_3\text{CH}_2\text{PO}_4\text{H}_2$).

4.2.1 Paper III

Following previous computational studies, lumiflavin (LF, Fig. 4.5) was used as a model of the FMN chromophore. Besides its photobiological relevance, there are also technical reasons why LF is a suitable test system for the relativistic TD-DFT calculations performed in **Paper III**. Most importantly, it has been shown that for this system the performance of TD-DFT^{118–120} is comparable to that of multiconfigurational methods like DFT-based multireference configuration interaction^{118,119} and CASPT2.¹²¹

Figure 4.5. Chemical structure of lumiflavin (LF). Reprinted from **Paper III** with permission from The Journal of Physical Chemistry A. Copyright 2015 American Chemical Society.

To identify the excited states relevant for the singlet and triplet photochemistry of LF, relativistic 4C-TD-DFT/B3LYP/aug-cc-pVDZ calculations were performed to obtain energies and oscillator strengths for transitions from the ground state to low-lying electronically excited states. From the results presented in Table 4.2, we first note that the transition with the largest oscillator strength ($f \sim 10^{-1}$) is the transition to the $^1(\pi_H\pi_L^*)$ state (i.e., S_1). Since the oscillator strengths for the transitions to the other excited states are at least two orders of magnitude smaller ($f \leq \sim 10^{-3}$), it seems very likely that $^1(\pi_H\pi_L^*)$ is the initial absorbing state. Furthermore, the calculations also predict that the lowest triplet state (i.e., T_1), wherein the chemical reaction with the cysteine residue occurs, is the $^3(\pi_H\pi_L^*)$ state. Finally, since the $^3(n_{N2}\pi_L^*)$ and $^3(\pi_{H-1}\pi_L^*)$ states lie between $^3(\pi_H\pi_L^*)$ and $^1(\pi_H\pi_L^*)$, it is possible that the relaxation of the absorbing state into the reactive state proceeds via both or one of these states.

Table 4.2. Calculated 4C-TD-DFT energies (ΔE , in eV) and oscillator strengths (f) for transitions from the electronic ground state to low-lying electronically excited states of LF.^a Reprinted from **Paper III** with permission from The Journal of Physical Chemistry A. Copyright 2015 American Chemical Society.

state	ΔE	f
$^3(\pi_H\pi_L^*)$	2.04	$7.306 \cdot 10^{-10}$
$^3(n_{N2}\pi_L^*)$	2.77	$1.339 \cdot 10^{-5}$
$^3(\pi_{H-1}\pi_L^*)$	2.80	$6.610 \cdot 10^{-9}$
$^1(\pi_H\pi_L^*)$	2.96	$1.930 \cdot 10^{-1}$
$^3(n_{O2}\pi_L^*)$	2.97	$5.358 \cdot 10^{-3}$
$^1(n_{O2}\pi_L^*)$	3.12	$2.053 \cdot 10^{-4}$
$^1(n_{N2}\pi_L^*)$	3.28	$3.719 \cdot 10^{-4}$
$^3(\pi_{H-2}\pi_L^*)$	3.36	$4.278 \cdot 10^{-8}$
$^3(\pi_H\pi_{L+1}^*)$	3.58	$9.614 \cdot 10^{-9}$

^aAll calculations carried out in the gas phase using B3LYP, the aug-cc-pVDZ basis set, and a ground-state geometry optimized in a non-relativistic framework.

Having found that the $^3(\pi_H\pi_L^*)$, $^3(n_{N2}\pi_L^*)$, $^3(\pi_{H-1}\pi_L^*)$ and $^1(\pi_H\pi_L^*)$ states are

likely to be involved in the singlet and triplet photochemistry of LF, we then continued by predicting which of the possible radiative and non-radiative pathways from the absorbing $^1(\pi_{\text{H}}\pi_{\text{L}}^*)$ state is the preferred one. This was done by calculating radiative and non-radiative rate constants for the transitions to the $^3(\pi_{\text{H}}\pi_{\text{L}}^*)$, $^3(n_{\text{N2}}\pi_{\text{L}}^*)$ and $^3(\pi_{\text{H-1}}\pi_{\text{L}}^*)$ states. As mentioned above, this was the first time that first-principles calculations are performed to estimate radiative rate constants (k_{rad}) for inter-excited phosphorescence events. These rate constants were obtained from 4C-TD-DFT calculations using the expression (in SI-units)

$$k_{\text{rad}} = \frac{1}{3\pi\epsilon_0 c^3 \hbar^4} \sum_k \left[(\Delta E^k)^3 \sum_{\alpha \in \{x,y,z\}} |M_{\alpha}^k|^2 \right], \quad (4.1)$$

where ϵ_0 is the vacuum permittivity, k indicates that the sum is over the three different sublevels of the final triplet state, and ΔE^k is transition energy and M_{α}^k the transition dipole moment (along the Cartesian axis α in the molecular frame) between the singlet state and the specific sublevel k of the triplet state.

The non-radiative rate constants (k_{nonrad}), in turn, were obtained using the expression (in SI-units)

$$k_{\text{nonrad}} = \frac{2\pi}{\hbar} \sum_k |H_{\text{SO}}^k|^2 \rho(E^k), \quad (4.2)$$

where H_{SO}^k is the spin-orbit coupling between non-relativistic states and $\rho(E^k)$ is the vibrational density of states for sublevel k of the final triplet state at the energy of the initial singlet state. $\rho(E^k)$ was estimated in three different ways to yield (through calculation of the spin-orbit coupling) rate constants denoted $k_{\text{nonrad}}^{\rho_0,1}$, $k_{\text{nonrad}}^{\rho_0,2}$ and $k_{\text{nonrad}}^{\text{EGL}}$.

Table 4.3. Calculated radiative (k_{rad}) and non-radiative (k_{nonrad}) rate constants for relaxation pathways from the $^1(\pi_{\text{H}}\pi_{\text{L}}^*)$ state of LF (in s^{-1}).^a Reprinted from **Paper III** with permission from The Journal of Physical Chemistry A. Copyright 2015 American Chemical Society.

final state	k_{rad}	$k_{\text{nonrad}}^{\rho_0,1}$	$k_{\text{nonrad}}^{\rho_0,2}$	$k_{\text{nonrad}}^{\text{EGL}}$
$^3(\pi_{\text{H}}\pi_{\text{L}}^*)$	$3 \cdot 10^0$	$2 \cdot 10^2$	$6 \cdot 10^3$	$9 \cdot 10^0$
$^3(n_{\text{N2}}\pi_{\text{L}}^*)$	$2 \cdot 10^2$	$4 \cdot 10^7$	$1 \cdot 10^9$	$8 \cdot 10^7$
$^3(\pi_{\text{H-1}}\pi_{\text{L}}^*)$	$7 \cdot 10^{-2}$	$2 \cdot 10^2$	$6 \cdot 10^3$	$6 \cdot 10^2$

^aAll calculations carried out in the gas phase using B3LYP, the aug-cc-pVDZ basis set, and a ground-state geometry optimized in a non-relativistic framework. k_{nonrad} values were estimated in three different ways.

From Table 4.3, it is observed that irrespective of which scheme is used to estimate the non-radiative rate constants ($k_{\text{nonrad}}^{\rho_0,1}$, $k_{\text{nonrad}}^{\rho_0,2}$ or $k_{\text{nonrad}}^{\text{EGL}}$), these pathways to the three triplet states are consistently more favorable than the radiative pathways. Indeed, except for the $k_{\text{nonrad}}^{\text{EGL}}$ value of 9 s^{-1} for the transition to $^3(\pi_{\text{H}}\pi_{\text{L}}^*)$, which is only three times larger than the corresponding k_{rad} value of 3 s^{-1} , the

three different estimates of the non-radiative rate constants are for each transition at least two orders of magnitude larger than the corresponding radiative rate constant. Hence, the calculations clearly indicate that the transition from the $^1(\pi_H\pi_L^*)$ state to the triplet manifold proceeds through a non-radiative mechanism. Comparing the k_{nonrad} values for the different triplet states, we note that the transition to the $^3(n_{N2}\pi_L^*)$ state (values of $4 \cdot 10^7 - 1 \cdot 10^9 \text{ s}^{-1}$) is expected to be much faster than those to the $^3(\pi_H\pi_L^*)$ ($9 \cdot 10^0 - 6 \cdot 10^3 \text{ s}^{-1}$) and $^3(\pi_{H-1}\pi_L^*)$ ($2 \cdot 10^2 - 6 \cdot 10^3 \text{ s}^{-1}$) states. Thus, the calculated rate constants suggest that the relaxation of the absorbing $^1(\pi_H\pi_L^*)$ state into the triplet manifold occurs non-radiatively via the $^3(n_{N2}\pi_L^*)$ state.

In summary, using LF as a prototypical example, **Paper III** showed, for the first time, that first-principles calculations can be performed to obtain radiative rate constants for phosphorescence events occurring between two different excited states. By comparing these rate constants with the corresponding non-radiative ones, it was found that for these particular transitions the radiative rate constants are typically two to five orders of magnitude smaller. Furthermore, it was also found that the relaxation between the absorbing and reactive states occurs via the $^3(n_{N2}\pi_L^*)$ state, which is formed in a non-radiative fashion.

4.3 Phytochromes

Figure 4.6. Structure of the biliverdin IX α bilin chromophore. Reprinted from **Paper V** with permission from John Wiley & Sons. Copyright 2013 Wiley Periodicals, Inc.

Phytochromes constitute a superfamily of photosensory proteins that control processes such as seed germination, flowering time and shade avoidance in plants, and phototaxis in bacteria. The red/far-red absorbing bilin chromophore (Fig. 4.6) is covalently attached through a thioether linkage to a cysteine residue of the apoprotein. Apart from the characteristic strong absorption of red light (the so-called Q band), UV-vis absorption spectra of bilin chromophores also exhibit a band in the blue region (the so-called Soret band).¹⁰¹

Figure 4.7. Photocycle of phytochromes. $h\nu$ and Δ indicate light and thermally-induced reaction pathways between the two forms, respectively.

The biological activity of phytochromes depends on the relative concentrations of a red-absorbing (Pr) and a far-red-absorbing (Pfr) form. In most phytochromes, the Pr form is favored in the dark-adapted inactive state, whereas the Pfr form is favored in the biologically active state.¹⁰¹ Absorption of red light by Pr triggers the conversion of Pr into Pfr. This conversion begins with a $Z \rightarrow E$ photoisomerization of the bilin chromophore and is followed by a number of thermal steps (Fig. 4.7). Although direct evidence is scarce, it is generally accepted that the photoisomerization involves a rotation of pyrrole ring D around the axis of the C15-C16 bond of the methine bridge between rings C and D. Support for such a mechanism comes from, for example, X-ray crystallographic data and studies using a variety of spectroscopic techniques such as resonance Raman and magic angle spin-

ning NMR.^{122–125} In the gas phase and in solution, on the other hand, it has been found that bilin chromophores have a tendency to rather photoisomerize at the C10-C11 bond.^{126–129} Indirectly, this finding indicates that interactions between the chromophore and the apoprotein play a decisive role not only to facilitate a fast reaction at the the C15-C16 bond, but also to prevent the reaction from taking place at the C10-C11 bond.

Figure 4.8. Structure of the chromophore binding domain of *Deinococcus radiodurans* bacteriophytochrome in the Pr form. Reprinted from **Paper V** with permission from John Wiley & Sons. Copyright 2013 Wiley Periodicals, Inc.

In **Paper VI**, we studied how phytochromes control the photochemical reactivity of their bilin chromophores. Specifically, by performing QM/MM calculations on *Deinococcus radiodurans* bacteriophytochrome (DrBphP, Fig. 4.8), we investigated how the protein discriminates between competing isomerization pathways of its biliverdin IX α (BV) bilin chromophore (Fig. 4.6). Although phytochromes play key roles in the regulation of many important processes in Nature, only a few computational studies have so far been devoted to these systems. Therefore, to identify a suitable way of modeling the photoactivation of DrBphP, two systematic studies (**Papers IV** and **V**) were first undertaken to assess the performance of QM and QM/MM methods in the modeling of spectral properties of phytochromes. In **Paper IV**, the performance of primarily TD-DFT but also some *ab initio* methods was assessed by computing UV-vis absorption and fluorescence spectra of four

synthetic bilin derivatives. In **Paper V**, in turn, we investigated how the choice of QM/MM methodology affects the resulting UV-vis absorption spectra of *DrBphP*.

4.3.1 Paper IV

By performing TD-DFT calculations on four sterically locked bilin chromophores (denoted *5Za15Ea*, *5Ea15Ea*, *5Zs15Za*, *15Za*) in aqueous solution, we investigated how the choice of density functional affects the resulting absorption and fluorescence spectra. The calculations were performed with a set of representative density functionals including BP86, BLYP and PBE (GGAs), τ -HCTH (a meta-GGA), B3LYP, BHHLYP, PBE0 (hybrid GGAs), M06-2X, M06-HF (hybrid meta-GGAs), LC-BLYP, LC- ω PBE, CAM-B3LYP and ω B97X-D (range-separated hybrid GGAs). While this set comprises some widely used standard functionals like B3LYP and PBE0, it also includes functionals that, at least in part, have been designed to provide a more accurate description of excited states such as CAM-B3LYP and LC-BLYP.

The structures of the chromophores studied are shown in Fig. 4.9. Absorption maxima were estimated as vertical $S_0 \rightarrow S_1$ excitation energies and fluorescence maxima as vertical $S_1 \rightarrow S_0$ emission energies. These approximations were, in a number of cases, validated by comparing $S_0 \rightarrow S_1$ excitation energies with the corresponding full absorption spectra obtained from a multitude of Gaussian-convoluted transition energies. Based on optimized geometries using B3LYP-PCM/SVP, absorption and fluorescence maxima were obtained at the TD-DFT/cc-pVDZ level of theory in combination with a state-specific PCM description of the water solvent. While non-equilibrium solvation was used for calculating absorption maxima, equilibrium solvation was used for fluorescence maxima.

For the absorption maxima, it was found that for all four chromophores, all density functionals give results that are blue-shifted relative to corresponding experimental values. Furthermore, it was also found that the pure GGAs (BP86, BLYP, PBE and τ -HCTH) exhibit better agreement with the experimental values than the hybrid GGAs (BHHLYP, B3LYP and PBE0), as well as than the hybrid meta-GGAs (M06-2X and M06-HF) and the range-separated hybrid GGAs (LC-BLYP, CAM-B3LYP and ω B97X-D). Specifically, while the pure GGA values deviate from the experimental values by ~ 0.15 eV, the values computed with the other functionals differ from the reference values by more than ~ 0.3 eV. While this observation is not in line with how pure GGAs typically perform,³² the same trend was actually observed in a QM study on the chemically related porphyrin molecule.⁸ For the fluorescence maxima, in turn, it was observed that also for these do the pure GGAs afford the best performance.

Although TD-DFT often yields accurate excitation energies for valence excited states, there are some well-known failures with this approach such as the treatment of states with double-excitation character. To investigate if such effects are important for bilin chromophores, complementary calculations were carried out with the CASPT2 method. By analyzing the CASSCF wave function underlying the CASPT2 calculations, we observed that such effects are of minor importance for the S_1 states.

Figure 4.9. Structures of the locked bilin chromophores considered ($R = \text{CH}_2\text{CH}_2\text{COOH}$). Reprinted from **Paper IV** with permission from Elsevier. Copyright 2013 Elsevier B.V.

To further assess the performance of the density functionals, we also compared the calculated Stokes shifts (difference between absorption and fluorescence maxima) with the corresponding experimental values. From this analysis, we found that also in this case do the pure GGAs (Stokes shifts in the 0.07 – 0.10 eV range) show the best agreement with the experimental values (0.04 – 0.07 eV), followed by B3LYP and PBE0 (0.12 – 0.16 eV), which perform slightly better than the other hybrid functionals (0.18 – 0.25 eV). Taken together, the results of **Paper IV** support the conclusion that the well-established trend that hybrid functionals perform better than pure functionals does *not* apply to bilin chromophores.

4.3.2 Paper V

As a natural continuation of **Paper IV**, **Paper V** focused on how the choice of density functional for the QM system affects the computed UV-vis absorption spectra of bilin chromophores in a protein environment. Specifically, using a two-layer ONIOM(QM:MM) approach, QM/MM calculations were performed to obtain the Q and Soret-band absorption maxima (based on a multitude of Gaussian-convoluted excitation energies) of DrBphP in the Pr form (Fig. 4.8). In the calculations, the full chromophore binding pocket (the BV chromophore and its neighboring amino-acid residues and water molecules) was relaxed. The QM system, consisting of the BV chromophore and the side chain of Cys24, was treated with TD-DFT, whereas the MM part was modeled with the AMBER force field. The computed Q and Soret-band absorption maxima are compared with the corresponding experimental values in Table 4.4.

Table 4.4. Calculated absorption maxima ($\lambda_{\max}$, in eV) and relative oscillator strengths (f_Q/f_{Soret}) of the Q and Soret bands of DrBphP using different density functionals for the QM system.^a Reprinted from **Paper V** with permission from John Wiley & Sons. Copyright 2013 Wiley Periodicals, Inc.

Functional	$\lambda_{\max}(\text{Q})$	$\lambda_{\max}(\text{Soret})$	f_Q/f_{Soret}
BP86	1.91	3.15	1.48
BLYP	1.90	3.13	1.48
B97D	1.91	3.16	1.49
τ -HCTH	1.91	3.17	1.49
B3LYP	2.05	3.20	1.65
PBE0	2.10	3.27	1.54
M06-HF	2.26	3.77	1.04
LC-BLYP	2.31	3.83	1.08
CAM-B3LYP	2.20	3.54	1.19
LC- ω PBE	2.29	3.78	1.08
Exp. ^b	1.77	3.26	—

^aAll QM/MM calculations based on the ONIOM(B3LYP/6-31G(d,p):AMBER)-optimized structure and carried out with the default QM system and the 6-31+G(d,p) basis set. ^bExperimental data from Ref. 130.

Focusing on the photochemically relevant Q band, we note that the results obtained with the pure GGAs (BP86, BLYP, B97D and τ -HCTH) are again in better agreement with the experimentally observed absorption maximum than the values obtained with the hybrid GGAs (B3LYP and PBE0). Specifically, while the pure GGAs deviate from the experimental value by less than 0.15 eV, the hybrid GGAs (B3LYP and PBE0) and also those functionals that have been designed to treat excited states (M06-HF, LC-BLYP, CAM-B3LYP and LC- ω PBE) blue-shift this peak by up to 0.5 eV.

Having carried out QM/MM calculations on the DrBphP protein with a number of density functionals ($\lambda_{\max}^{(3)}$ in Table 4.5), we also investigated how well these functionals are able to capture the response of the chromophore to the electric field generated by the protein environment. To this end, complementary TD-DFT calculations were performed in the gas phase determining the Q-band absorption maximum of the bare BV chromophore extracted from the QM/MM optimized structure ($\lambda_{\max}^{(2)}$ in Table 4.5). In the absence of an experimental Q-band absorption maximum in the gas phase, the reference value for these calculations was instead obtained from high-level *ab initio* calculations. Based on the estimated value in the gas phase (1.65 eV) and the experimental value of DrBphP (1.77 eV), it was predicted that the protein environment blue shifts the intrinsic gas-phase absorption by 0.12 eV ($\lambda_{\max}^{(2)} \rightarrow \lambda_{\max}^{(3)}$ in Table 4.5). To solidify this conclusion, supplementary QM and QM/MM calculations were performed at the CIS(D)/6-31G(d,p) level of theory, yielding a shift of 0.14 eV. This investigation is summarized in Table 4.5.

Interestingly, despite that the pure GGAs show better agreement with the experimental DrBphP Q-band maximum, it is actually the $\lambda_{\max}^{(2)} \rightarrow \lambda_{\max}^{(3)}$ values obtained with the hybrid functionals that best reproduce the reference value of 0.12 eV. Specifically, while the hybrid functionals predict blue shifts of 0.08 – 0.11 eV and also match the CIS(D) value of 0.14 eV, the pure GGAs give red shifts of 0.03 – 0.04 eV.

Besides a spectral shift induced by the electrostatic environment, the protein also perturbs the geometry of the chromophore, which may further alter the absorption. To quantify this geometric effect, complementary high-level *ab initio* calculations were first carried out to obtain a reference value pertaining to a relaxed chromophore geometry in isolated conditions ($\lambda_{\max}^{(1)}$ in Table 4.5). By comparing this value with the corresponding value obtained for the QM/MM geometry, it was predicted that the protein-induced change in geometry blue shifts the absorption by 0.10 eV ($\lambda_{\max}^{(1)} \rightarrow \lambda_{\max}^{(2)}$ in Table 4.5), whereby the total net effect of the protein environment on the Q-band absorption is approximately 0.12 + 0.10 $\approx$ 0.2 eV ($\lambda_{\max}^{(1)} \rightarrow \lambda_{\max}^{(3)}$ in Table 4.5). From Table 4.5, it is observed that all functionals associate the geometric perturbation ($\lambda_{\max}^{(1)} \rightarrow \lambda_{\max}^{(2)}$) with a blue shift of 0.09 – 0.17 eV, which is in good agreement with the reference value of 0.10 eV. For the total protein-induced shift ($\lambda_{\max}^{(1)} \rightarrow \lambda_{\max}^{(3)}$), finally, it is found that the hybrid functionals (0.19 – 0.28 eV) give the best performance relative to the reference value of 0.2 eV.

To summarize the results of **Paper V**, we conclude that while the pure GGAs

Table 4.5. Calculated absorption maxima ($\lambda_{\max}$, in eV) of the Q band of the BV chromophore in gas phase and protein environments^a Reprinted from **Paper V** with permission from John Wiley & Sons. Copyright 2013 Wiley Periodicals, Inc.

Env ^b →	$\lambda_{\max}^{(1)}$ Gas	$\lambda_{\max}^{(2)}$ Gas	$\lambda_{\max}^{(3)}$ Protein	$\lambda_{\max}^{(1)}$ →	$\lambda_{\max}^{(1)}$ →	$\lambda_{\max}^{(2)}$ →
Geo ^c →	Gas	Protein	Protein	$\lambda_{\max}^{(2)}$	$\lambda_{\max}^{(3)}$	$\lambda_{\max}^{(3)}$
Functional						
BP86	1.84	1.94	1.91	0.10	0.07	−0.03
BLYP	1.83	1.93	1.90	0.10	0.07	−0.03
B97D	1.84	1.95	1.91	0.11	0.07	−0.04
τ -HCTH	1.85	1.95	1.91	0.10	0.06	−0.04
B3LYP	1.86	1.96	2.05	0.10	0.19	0.09
PBE0	1.91	2.00	2.10	0.09	0.19	0.10
M06-HF	2.00	2.17	2.26	0.17	0.26	0.09
LC-BLYP	2.03	2.20	2.31	0.17	0.28	0.11
CAM-B3LYP	1.99	2.12	2.20	0.13	0.21	0.08
LC- ω PBE	2.03	2.18	2.29	0.15	0.26	0.11
CIS(D) ^d	1.94	2.05	2.19	0.11	0.25	0.14
Reference value	1.55 ^e	1.65 ^e	1.77 ^f	0.10	0.22	0.12

^aUnless otherwise noted, all calculations carried out with the 6-31+G(d,p) basis set. ^bEnvironment = Gas: QM calculations carried out in the gas phase. Environment = Protein: QM/MM calculations carried out with the default QM system. ^cGeometry = Gas: Calculations based on the B3LYP/6-31G(d,p) optimized BV structure. Geometry = Protein: Calculations based on the QM/MM optimized structure. ^dCIS(D) absorption maxima calculated with the 6-31G(d,p) basis set and obtained as $S_0 \rightarrow S_1$ excitation energies. ^eSAC-CI-extrapolated value. ^fExperimental data from Ref. 130.

better reproduce the experimental Q-band absorption maximum of *DrBphP*, the hybrid functionals account more accurately for spectral shifts induced by the protein environment. These results are very useful for the selection of a suitable density functional for studying the phytochrome photoisomerization, which was the topic of **Paper VI**.

4.3.3 Paper VI

Having assessed the accuracy of different density functionals for modeling the absorption and fluorescence spectra of bilin chromophores in solvent and protein environments in **Papers IV** and **V**, **Paper VI** is concerned with the light-induced chromophore reaction that initiates the conversion of Pr into Pfr.

QM/MM calculations were first carried out to assess the competition between possible isomerization pathways of the BV chromophore bound to *DrBphP* (Fig. 4.8). The QM/MM calculations were performed using the scheme employed in **Paper V**. The QM region was described using the ω B97X-D functional, and

the MM region was treated with the AMBER force field.

Although the photoisomerization can in principle occur at any of the C4-C5, C5-C6, C9-C10, C10-C11, C14-C15 and C15-C16 bonds of the BV chromophore (atom numbering is shown Fig. 4.6), it was found from an initial Franck-Condon-relaxation-based analysis that only the C4-C5, C9-C10 and C15-C16 bonds are conceivable sites for the photoisomerization.

Figure 4.10. ω B97X-D/AMBER excited-state potential energy curves for the initial stages of photoisomerization at the C4-C5, C9-C10, C10-C11 and C15-C16 bonds of BV in Pr DrBphP. Energies (ΔE) and dihedral angles ($|\Delta\omega|$) are given relative to their ground-state Pr values.

To determine which of the C4-C5, C9-C10 and C15-C16 bonds is the preferred one for photoisomerization, we then mapped the regions of the excited-state potential energy surface corresponding to up to $\pm 50^\circ$ torsional motion along the C3-C4-C5-C6, C8-C9-C10-C11 and C14-C15-C16-C17 dihedral coordinates, respectively. The results from these calculations are summarized in Fig. 4.10.

Notably, it is observed that the energy barrier to complete the $\pm 50^\circ$ rotation at the C15-C16 bond is about 5 kcal mol $^{-1}$, which is much smaller than the 14 and 19 kcal mol $^{-1}$ required by the reactions at the C4-C5 and C9-C10 bonds, respectively. This is a clear indication that the C15-C16 bond is the preferred photoisomerization site. While this is in line with the widely accepted view that the activation of phytochromes is triggered by a rotation of ring D around the axis of the C15-C16 bond, we then proceeded to investigate why this is so.

One possible explanation is that the chromophore reaction pathways offered by the apoprotein for the different isomerizations are such that the reaction at the C15-C16 bond is intrinsically more favorable for the BV chromophore. To test this hypothesis, QM excited-state singlepoint calculations were carried out on the QM/MM chromophore geometries along the C4-C5, C9-C10 and C15-C16 reaction pathways. The resulting energies, denoted E_{BV} , are presented in Fig. 4.11 (top). Notably, however, the C15-C16 intrinsic energy profile is very similar to those for

photoisomerizations at C4-C5 and C9-C10. This means that another mechanism must be at play.

Figure 4.11. (Top) ω B97X-D excited-state potential energy curves for the bare chromophore during the initial stages of photoisomerization at the C4-C5, C9-C10 and C15-C16 bonds of BV in Pr DrBphP. (Bottom) AMBER E_{vdW} curves for the initial stages of photoisomerization at the C4-C5, C9-C10 and C15-C16 bonds of BV in Pr DrBphP. Energies (ΔE) and dihedral angles ($|\Delta\omega|$) are given relative to their ground-state Pr values.

Another possible explanation is that steric repulsion between the chromophore and the apoprotein is smaller for the photoisomerization at the C15-C16 bond than at the C4-C5 and C9-C10 bonds. In this regard, it is worth noticing that, prior to photoexcitation, available Pr crystal structures show that rings A-C are more tightly packed by the protein than ring D.^{122,123} To test the hypothesis that steric effects are critical for the photoactivation, MM singlepoint calculations

were performed (with the AMBER force field) to obtain the van der Waals energy (E_{vdW}) of the full system for each geometry along the C4-C5, C9-C10 and C15-C16 reaction pathways. Interestingly, Fig. 4.11 (bottom) shows that the E_{vdW} cost to complete the 50° rotation is ~ 6 and ~ 10 kcal mol $^{-1}$ larger at the C4-C5 and C9-C10 sites, respectively, than at the C15-C16 site. Hence, steric effects are the major parts of the ~ 9 (C4-C5) and ~ 14 (C9-C10) kcal mol $^{-1}$ by which the corresponding reaction barriers in Fig. 4.10 are larger than the barrier at C15-C16. Thus, the key finding from **Paper VI** is that steric effects play a decisive role for how phytochromes are activated by light.

CHAPTER 5

Concluding Remarks

This thesis has been devoted to computational studies for improving the understanding of photobiological keto-enol reactions and chromophores. By performing QM and QM/MM calculations, photophysical and photochemical properties of the AXT, OxyLH₂, LF and bilin chromophores have been investigated. The key results can be summarized as follows:

- The dramatic change in color from deep-blue to red upon cooking of live lobsters seems to originate from a conversion of AXT from its enolate form (**1**⁻) to the neutral keto form (**1**).
- The phenolate-keto-OxyLH⁻ form is the intrinsically preferred excited-state tautomer of the OxyLH₂ light emitter of fireflies.
- First-principles methods can be used to obtain rate constants for radiative transitions from one excited state to another excited state of different spin symmetry, as demonstrated by 4C relativistic calculations on LF.
- The calculation of absorption and fluorescence spectra of bilin chromophores and phytochromes are well-amenable to DFT-based QM and QM/MM methods, although it is difficult to find a density functional that is capable of both reproducing experimental spectra in an absolute sense *and* accurately predict spectral shifts.
- Steric interactions between the bilin chromophore and the surrounding protein are a key discriminator between the different possible reaction channels for the photoactivation of phytochromes.

Needless to say, many unresolved questions remain in these fields of research. As for AXT, it would be of interest to investigate how specific chromophore-protein

interactions and long-range electrostatic bulk effects influence the absorption maxima of the various forms of the chromophore inside the binding pocket of crustacyanins. While a previous computational study has shown that the neutral keto form (**1**) is rather insensitive to short-range chromophore-protein interactions,¹³¹ preliminary calculations reported in the thesis indicate that this is not the case for the enolate form (**1**⁻).

For OxyLH₂, a natural next step would be to examine the excited-state equilibria of the chromophore inside the active pocket of firefly luciferase. Although the intrinsic stabilities of the various forms of OxyLH₂ studied in this work do give useful information on the *in vivo* excited-state equilibria, the considerable challenge to calculate, e.g., excited-state pK_a values in a protein environment is a worthwhile endeavor for future studies.

Having demonstrated that the well-established ability of first-principles methods to describe non-radiative channels for transitions from one excited state to another excited state of different spin multiplicity, can be supplemented by the ability to also describe the corresponding radiative channels, it would be of interest to quantitatively assess the competition between radiative and non-radiative channels for organic molecules exhibiting phosphorescence from high-lying excited states.

As for future studies of phytochromes, it would be intriguing to assess whether steric effects are also important for the photochemical deactivation of the Pfr form into the Pr form, as well as for the thermal steps of the photocycle. Methodologically, it is a natural goal to also simulate the molecular dynamics of the full photocycle.

Finally, it is my hope that the results presented in this thesis have not only provided a better understanding of the actual systems studied, but will also be useful for future studies of related systems.

Bibliography

- [1] McNaught, A. D., Wilkinson, A., Eds. *IUPAC. Compendium of Chemical Terminology*, 2nd ed.; Blackwell Scientific Publications: Oxford, 1997.
- [2] Pariser, R.; Parr, R. A Semi-Empirical Theory of the Electronic Spectra and Electronic Structure of Complex Unsaturated Molecules. I. *J. Chem. Phys.* **1953**, *21*, 466.
- [3] Pople, J. A. Electron Interaction in Unsaturated Hydrocarbons. *Trans. Faraday Soc.* **1953**, *49*, 1375.
- [4] Roos, B. O.; Taylor, P. R.; Siegbahn, P. E. M. A Complete Active Space SCF Method (CASSCF) Using a Density Matrix Formulated Super-CI Approach. *Chem. Phys.* **1980**, *48*, 157.
- [5] Bernardi, F.; De, S.; Olivucci, M.; Robb, M. A. Mechanism of Ground-State-Forbidden Photochemical Pericyclic Reactions: Evidence for Real Conical Intersections. *J. Am. Chem. Soc.* **1990**, *112*, 1737.
- [6] Foresman, J. B.; Head-Gordon, M.; Pople, J. A.; Frisch, M. J. Toward a Systematic Molecular Orbital Theory for Excited States. *J. Chem. Phys.* **1992**, *96*, 135.
- [7] Runge, E.; Gross, E. K. U. Density-Functional Theory for Time-Dependent Systems. *Phys. Rev. Lett.* **1984**, *52*, 997.
- [8] Bauernschmitt, R.; Ahlrichs, R. Treatment of Electronic Excitations within the Adiabatic Approximation of Time Dependent Density Functional Theory. *Chem. Phys. Lett.* **1996**, *256*, 454.
- [9] Petersilka, M.; Gossmann, U. J.; Gross, E. K. U. Excitation Energies from Time-Dependent Density-Functional Theory. *Phys. Rev. Lett.* **1996**, *76*, 1212.

- [10] Casida, M. E.; Jamorski, C.; Casida, K. C.; Salahub, D. R. Molecular Excitation Energies to High-Lying Bound States from Time-Dependent Density-Functional Response Theory: Characterization and Correction of the Time-Dependent Local Density Approximation Ionization Threshold. *J. Chem. Phys.* **1998**, *108*, 4439.
- [11] Stratmann, R. E.; Scuseria, G. E.; Frisch, M. J. An Efficient Implementation of Time-Dependent Density-Functional Theory for the Calculation of Excitation Energies of Large Molecules. *J. Chem. Phys.* **1996**, *109*, 8218.
- [12] Warshel, A.; Levitt, M. Theoretical Studies of Enzymic Reactions: Dielectric, Electrostatic and Steric Stabilization of the Carbonium Ion in the Reaction of Lysozyme. *J. Mol. Biol.* **1976**, *103*, 227.
- [13] Frutos, L. M.; Andruniów, T.; Santoro, F.; Ferré, N.; Olivucci, M. Tracking the Excited-State Time Evolution of the Visual Pigment with Multiconfigurational Quantum Chemistry. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 7764.
- [14] Falklöf, O. *Photochemical Properties of Phytochrome and Firefly Luciferase Chromophores: A Theoretical Study*; Licentiate Thesis. Linköping University, Sweden, 2014.
- [15] Jensen, F. *Introduction to Computational Chemistry*, 2nd ed.; John Wiley & Sons, 2007.
- [16] Andersson, K.; Malmqvist, P.-Å.; Roos, B. O. Second-Order Perturbation Theory with a Complete Active Space Self-Consistent Field Reference Function. *J. Chem. Phys.* **1992**, *96*, 1218.
- [17] Hohenberg, P.; Kohn, W. Inhomogeneous Electron Gas. *Phys. Rev.* **1964**, *136*, B864.
- [18] Nordholm, S. Analysis of Covalent Bonding by Nonergodic Thomas-Fermi Theory. *J. Chem. Phys.* **1987**, *86*, 363.
- [19] von Weizsäcker, C. F. Zur Theorie der Kernmassen. *Z. Phys.* **1935**, *96*, 431.
- [20] Kohn, W.; Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.* **1965**, *140*, A1133.
- [21] Perdew, J.; Schmidt, K. Jacob's Ladder of Density Functional Approximations for the Exchange-Correlation Energy. *AIP Conf. Proc.* **2001**, *577*, 1.
- [22] Casida, M. E. In *Recent Advances in Density Functional Methods Part I*; Chong, D. P., Ed.; World Scientific, 1995.
- [23] Van Caillie, C.; Amos, R. D. Geometric Derivatives of Excitation Energies Using SCF and DFT. *Chem. Phys. Lett.* **1999**, *308*, 249.

- [24] Van Caillie, C.; Amos, R. D. Geometric Derivatives of Density Functional Theory Excitation Energies Using Gradient-Corrected Functionals. *Chem. Phys. Lett.* **2000**, *317*, 159.
- [25] Furche, F.; Ahlrichs, R. Adiabatic Time-Dependent Density Functional Methods for Excited State Properties. *J. Chem. Phys.* **2002**, *117*, 7433.
- [26] Scalmani, G.; Frisch, M. J.; Mennucci, B.; Tomasi, J.; Cammi, R.; Barone, V. Geometries and Properties of Excited States in the Gas Phase and in Solution: Theory and Application of a Time-Dependent Density Functional Theory Polarizable Continuum Model. *J. Chem. Phys.* **2006**, *124*, 094107.
- [27] Caricato, M.; Mennucci, B.; Tomasi, J.; Ingrosso, F.; Cammi, R.; Corni, S.; Scalmani, G. Formation and Relaxation of Excited States in Solution: A New Time Dependent Polarizable Continuum Model Based on Time Dependent Density Functional Theory. *J. Chem. Phys.* **2006**, *124*, 124520.
- [28] Iikura, H.; T., T.; Yanai, T.; Hirao, K. Long-Range Correction Scheme for Generalized-Gradient-Approximation Exchange Functionals. *J. Chem. Phys.* **2001**, *115*, 3540.
- [29] Yanai, T.; Tew, D. P.; Handy, N. C. A New Hybrid Exchange-Correlation Functional Using the Coulomb-Attenuating Method (CAM-B3LYP). *Chem. Phys. Lett.* **2004**, *393*, 51.
- [30] Chai, J.-D.; Head-Gordon, M. Long-Range Corrected Hybrid Density Functionals with Damped Atom-Atom Dispersion Corrections. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615.
- [31] Dreuw, A.; Head-Gordon, M. Single-Reference Ab Initio Methods for the Calculation of Excited States of Large Molecules. *Chem. Rev.* **2005**, *105*, 4009.
- [32] Jacquemin, D.; Wathelet, V.; Perpète, E.; Adamo, C. Extensive TD-DFT Benchmark: Singlet-Excited States of Organic Molecules. *J. Chem. Theory Comput.* **2009**, *5*, 2420.
- [33] Head-Gordon, M.; Rico, J. R.; Oumi, M.; Lee, T. J. A Doubles Correction to Electronic Excited States from Configuration Interaction in the Space of Single Substitutions. *Chem. Phys. Lett.* **1994**, *219*, 21.
- [34] Nakatsuji, H. Cluster Expansion of the Wave Function. Excited States. *Chem. Phys. Lett.* **1978**, *59*, 362.
- [35] Stanton, J. F.; Bartlett, R. J. The Equation of Motion Coupled-Cluster Method. A Systematic Biorthogonal Approach to Molecular Excitation Energies, Transition Probabilities, and Excited State Properties. *J. Chem. Phys.* **1993**, *98*, 7029.

- [36] Christiansen, O.; Koch, H.; Jørgensen, P. The Second-Order Approximate Coupled Cluster Singles and Doubles Model CC2. *Chem. Phys. Lett.* **1995**, *243*, 409.
- [37] Dyall, K. G.; Fægri, K., Jr. *Introduction to Relativistic Quantum Chemistry*; Oxford University Press, 2007.
- [38] Liu, W. Ideas of Relativistic Quantum Chemistry. *Mol. Phys.* **2010**, *108*, 1679.
- [39] Saue, T. Relativistic Hamiltonians for Chemistry: A Primer. *ChemPhysChem* **2011**, *12*, 3077.
- [40] Widmark, P.-O., Ed. *European Summerschool in Quantum Chemistry*; 2011.
- [41] Pyykkö, P. The Physics behind Chemistry and the Periodic Table. *Chem. Rev.* **2012**, *112*, 371.
- [42] Liu, W.; Hong, G.; Dai, D.; Li, L.; Dolg, M. The Beijing Four-Component Density Functional Program Package (BDF) and Its Application to EuO, EuS, YbO and YbS. *Theor. Chem. Acc.* **1997**, *96*, 75.
- [43] Liu, W.; Wang, F.; Li, L. The Beijing Density Functional (BDF) Program Package: Methodologies and Applications. *J. Theor. Comput. Chem.* **2003**, *2*, 257.
- [44] Gao, J.; Liu, W.; Song, B.; Liu, C. Time-Dependent Four-Component Relativistic Density Functional Theory for Excitation Energies. *J. Chem. Phys.* **2004**, *121*, 6658.
- [45] Gao, J.; Zou, W.; Liu, W.; Xiao, Y.; Peng, D.; Song, B.; Liu, C. Time-Dependent Four-Component Relativistic Density-Functional Theory for Excitation Energies. II. The Exchange-Correlation Kernel. *J. Chem. Phys.* **2005**, *123*, 054102.
- [46] Bast, R.; Jensen, H. J. Aa.; Saue, T. Relativistic Adiabatic Time-Dependent Density Functional Theory Using Hybrid Functionals and Noncollinear Spin Magnetization. *Int. J. Quant. Chem.* **2009**, *109*, 2091.
- [47] Bast, R.; Jensen, H. J. Aa.; Saue, T.; Visscher, L.; et. al., *DIRAC, a Relativistic Ab Initio Electronic Structure Program, Release DIRAC11*; 2011.
- [48] Stanton, R. E.; Havriliak, S. Kinetic Balance: A Partial Solution to the Problem of Variational Safety in Dirac Calculations. *J. Chem. Phys.* **1984**, *81*, 1910.
- [49] Visscher, L. Approximate Molecular Relativistic Dirac-Coulomb Calculations Using a Simple Coulombic Correction. *Theor. Chem. Acc.* **1997**, *98*, 68.

- [50] Hess, B. A. Relativistic Electronic-Structure Calculations Employing a Two-Component No-Pair Formalism with External-Field Projection Operators. *Phys. Rev. A* **1986**, *33*, 3742.
- [51] van Lenthe, E.; Baerends, E. J.; Snijders, J. G. Relativistic Regular Two-Component Hamiltonians. *J. Chem. Phys.* **1993**, *99*, 4597.
- [52] Marian, C. M. Spin-Orbit Coupling and Intersystem Crossing in Molecules. *WIREs Comput. Mol. Sci.* **2012**, *2*, 187.
- [53] Dyall, K. G.; Taylor, P. R.; Fægri, K., Jr.; Partridge, H. All-Electron Molecular Dirac-Hartree-Fock Calculations: The Group IV Tetrahydrides CH₄, SiH₄, GeH₄, and PbH₄. *J. Chem. Phys.* **95**, 2583.
- [54] Cossi, M.; Barone, V. Time-Dependent Density Functional Theory for Molecules in Liquid Solutions. *J. Chem. Phys.* **2001**, *115*, 4708.
- [55] Improta, R.; Barone, V.; Scalmani, G.; Frisch, M. J. A State-Specific Polarizable Continuum Model Time Dependent Density Functional Theory Method for Excited State Calculations in Solution. *J. Chem. Phys.* **2006**, *125*, 054103.
- [56] Senn, H. M.; Thiel, W. QM/MM Methods for Biomolecular Systems. *Angew. Chem. Int. Ed.* **2009**, *48*, 1198.
- [57] Dapprich, S.; Komáromi, I.; Byun, K. S.; Morokuma, K.; Frisch, M. J. A New ONIOM Implementation in Gaussian98. Part I. The Calculation of Energies, Gradients, Vibrational Frequencies and Electric Field Derivatives. *J. Mol. Struct. (THEOCHEM)* **1999**, *462*, 1.
- [58] Chung, L. W.; Sameera, W. M. C.; Ramozzi, R.; Page, A. J.; Hatanaka, M.; Petrova, G. P.; Harris, T. V.; Li, X.; Li, Z.; Ke, Z.; Liu, F.; Li, H.-B.; Ding, L.; Morokuma, K. The ONIOM Method and Its Applications. *Chem. Rev.* **2015**, *115*, 5678.
- [59] Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M., Jr.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179.
- [60] Vreven, T.; Byun, K. S.; Komáromi, I.; Dapprich, S.; Montgomery, J. A., Jr.; Morokuma, K.; Frisch, M. J. Combining Quantum Mechanics Methods with Molecular Mechanics Methods in ONIOM. *J. Chem. Theory Comput.* **2006**, *2*, 815.
- [61] Wanko, M.; Hoffmann, M.; Strodel, P.; Koslowski, A.; Thiel, W.; Neese, F.; Frauenheim, T.; Elstner, M. Calculating Absorption Shifts for Retinal Proteins: Computational Challenges. *J. Phys. Chem. B* **2005**, *109*, 3606.

- [62] Tully, J. C.; Preston, R. K. Trajectory Surface Hopping Approach to Nonadiabatic Molecular Collisions: The Reaction of H^+ with D_2 . *J. Chem. Phys.* **1971**, *55*, 562.
- [63] Wang, W.; Hellinga, H. W.; Beese, L. S. Structural Evidence for the Rare Tautomer Hypothesis of Spontaneous Mutagenesis. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 17644.
- [64] Poelarends, G. J.; Veetil, V. P.; Whitman, C. P. The Chemical Versatility of the $\beta - \alpha - \beta$ Fold: Catalytic Promiscuity and Divergent Evolution in the Tautomerase Superfamily. *Cell. Mol. Life Sci.* **2008**, *65*, 3606.
- [65] Clayden, J.; Greeves, N.; Warren, S.; Wothers, P. *Organic Chemistry*; Oxford University Press, 2001.
- [66] Carey, F. A.; Sundberg, R. J. *Advanced Organic Chemistry Part A: Structure and Mechanisms*, 5th ed.; Springer, 2007.
- [67] McMurry, J. *Organic Chemistry*, 6th ed.; Thomson Learning: London, 2004.
- [68] Chiang, Y.; Kresge, A. J.; Tang, Y. S. pK_a and Keto-Enol Equilibrium Constant of Acetone in Aqueous Solution. *J. Am. Chem. Soc.* **1984**, *106*, 460.
- [69] Chiang, Y.; Hojatti, M.; Kresge, A. J.; Schepp, N. P.; Wirz, J. Vinyl Alcohol: Generation and Decay Kinetics in Aqueous Solution and Determination of the Tautomerization Equilibrium Constant and Acid Dissociation Constants of the Aldehyde and Enol Forms. *J. Am. Chem. Soc.* **1987**, *109*, 4000.
- [70] Keeffe, J. R.; Kresge, A. J.; Schepp, N. P. Keto-Enol Equilibrium Constants of Simple Monofunctional Aldehydes and Ketones in Aqueous Solution. *J. Am. Chem. Soc.* **1990**, *112*, 4862.
- [71] Spencer, J. N.; Holmboe, E. S.; Kirshenbaum, M. R.; Firth, D. W.; Pinto, P. B. Solvent Effects on the Tautomeric Equilibrium of 2,4-Pentanedione. *Can. J. Chem.* **1982**, *60*, 1178.
- [72] Yang, N. C.; Rivas, C. A New Photochemical Primary Process, the Photochemical Enolization of *o*-Substituted Benzophenones. *J. Am. Chem. Soc.* **1961**, *83*, 2213.
- [73] Sammes, P. G. Photoenolisation. *Tetrahedron* **1976**, *32*, 405.
- [74] Hart, H. Simple Enols. *Chem. Rev.* **1979**, *79*, 515.
- [75] Markov, P. Light-Induced Tautomerism of β -Dicarbonyl Compounds. *Chem. Soc. Rev.* **1984**, *13*, 69.
- [76] Yankov, P.; Saltiel, S.; Petkov, I. Photoketonization of Dibenzoylmethane in Polar Solvents. *Chem. Phys. Lett.* **1986**, *128*, 517.

- [77] Wan, P.; Budac, D.; Krogh, E. Excited State Carbon Acids: Base Catalysed Photoketonization of Dibenzosuberanol to Dibenzosuberoone via Initial C-H Bond Heterolysis from S₁. *J. Chem. Soc., Chem. Commun.* **1990**, 255.
- [78] Fita, P.; Luzina, E.; Dziembowska, T.; Kopeć, D.; Piątkowski, P.; Radzewicz, C.; Grabowska, A. Keto-Enol Tautomerism of Two Structurally Related Schiff Bases: Direct and Indirect Way of Creation of the Excited Keto Tautomer. *Chem. Phys. Lett.* **2005**, 416, 305.
- [79] Solntsev, K.; Laptinok, S.; Naumov, P. Photoinduced Dynamics of Oxy-luciferin Analogues: Unusual Enol "Super"photoacidity and Evidence for Keto-Enol Isomerization. *J. Am. Chem. Soc.* **2012**, 134, 16452.
- [80] Dellisanti, C. D.; Spinelli, S.; Cambillau, C.; Findlay, J. B. C.; Zagalsky, P. F.; Finet, S.; Receveur-Bréchet, V. Quaternary Structure of Alpha-Crustacyanin from Lobster as Seen by Small-Angle X-Ray Scattering. *FEBS Lett.* **2003**, 544, 189.
- [81] Buchwald, M.; Jencks, W. P. Optical Properties of Astaxanthin Solutions and Aggregates. *Biochemistry* **1968**, 7, 834.
- [82] Buchwald, M.; Jencks, W. P. Properties of the Crustacyanins and the Yellow Lobster Shell Pigment. *Biochemistry* **1968**, 7, 844.
- [83] Ilagan, R. P.; Christensen, R. L.; Chapp, T. W.; Gibson, G. N.; Pascher, T.; Polívka, T.; Frank, H. A. Femtosecond Time-Resolved Absorption Spectroscopy of Astaxanthin in Solution and in α -Crustacyanin. *J. Phys. Chem. A* **2005**, 109, 3120.
- [84] Cianci, M.; Rizkallah, P. J.; Olczak, A.; Raftery, J.; Chayen, N. E.; Zagalsky, P. F.; Helliwell, J. R. The Molecular Basis of the Coloration Mechanism in Lobster Shell: β -Crustacyanin at 3.2-Å Resolution. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, 99, 9795.
- [85] van Wijk, A. A. C.; Spaans, A.; Uzunbajakava, N.; Otto, C.; de Groot, H. J. M.; Lugtenburg, J.; Buda, F. Spectroscopy and Quantum Chemical Modeling Reveal a Predominant Contribution of Excitonic Interactions to the Bathochromic Shift in α -Crustacyanin, the Blue Carotenoprotein in the Carapace of the Lobster *Homarus gammarus*. *J. Am. Chem. Soc.* **2005**, 127, 1438.
- [86] Strambi, A.; Durbeej, B. Excited-State Modeling of the Astaxanthin Dimer Predicts a Minor Contribution from Exciton Coupling to the Bathochromic Shift in Crustacyanin. *J. Phys. Chem. B* **2009**, 113, 5311.
- [87] Christensson, N.; Žídek, K.; Magdaong, N. C. M.; LaFountain, A. M.; Frank, H. A.; Zigmantas, D. Origin of the Bathochromic Shift of Astaxanthin in Lobster Protein: 2D Electronic Spectroscopy Investigation of β -Crustacyanin. *J. Phys. Chem. B* **2013**, 117, 11209.

- [88] Kuhn, R.; Sörensen, N. A. Über die Farbstoffe des Hummers (*Astacus gammarus* L.). *Angew. Chem.* **1938**, *51*, 465.
- [89] Khan, M. M. T.; Martell, A. E. Metal Ion and Metal Chelate Catalyzed Oxidation of Ascorbic Acid by Molecular Oxygen. II. Cupric and Ferric Chelate Catalyzed Oxidation. *J. Am. Chem. Soc.* **1967**, *89*, 7104.
- [90] Abe, Y.; Horii, H.; Taniguchi, S.; Kamai, K.; Takagi, M. The Autoxidation of 2,3-Dihydroxy-2-propenal (Triose Reductone). The Effects of the pH on the Rate Constants. *Bull. Chem. Soc. Jpn.* **1983**, *56*, 467.
- [91] Fraga, H. Firefly Luminescence: A Historical Perspective and Recent Developments. *Photochem. Photobiol. Sci.* **2008**, *7*, 146.
- [92] Marques, S. M.; Esteves da Silva, J. C. G., Firefly Bioluminescence: A Mechanistic Approach of Luciferase Catalyzed Reactions. *IUBMB Life* **2009**, *61*, 6.
- [93] Naumov, P.; Ozawa, Y.; Ohkubo, K.; Fukuzumi, S. Structure and Spectroscopy of Oxyluciferin, the Light Emitter of the Firefly Bioluminescence. *J. Am. Chem. Soc.* **2009**, *131*, 11590.
- [94] Rebarz, M.; Kukovec, B.-M.; Maltsev, O.; Ruckebusch, C.; Hintermann, L.; Naumov, P.; Sliwa, M. Deciphering the Protonation and Tautomeric Equilibria of Firefly Oxyluciferin by Molecular Engineering and Multivariate Curve Resolution. *Chem. Sci.* **2013**, *4*, 3803.
- [95] Nakatani, N.; Hasegawa, J.-y.; Nakatsuji, H. Red Light in Chemiluminescence and Yellow-Green Light in Bioluminescence: Color-Tuning Mechanism of Firefly, *Photinus pyralis*, Studied by the Symmetry-Adapted Cluster-Configuration Interaction Method. *J. Am. Chem. Soc.* **2007**, *129*, 8756.
- [96] Chen, S.-F.; Liu, Y.-J.; Navizet, I.; Ferré, N.; Fang, W.-H.; Lindh, R. Systematic Theoretical Investigation on the Light Emitter of Firefly. *J. Chem. Theory Comput.* **2011**, *7*, 798.
- [97] Song, C.-I.; Rhee, Y. Dynamics on the Electronically Excited State Surface of the Bioluminescent Firefly Luciferase-Oxyluciferin System. *J. Am. Chem. Soc.* **2011**, *133*, 12040.
- [98] Branchini, B.; Murtiashaw, M.; Magyar, R.; Portier, N.; Ruggiero, M.; Stroth, J. Yellow-Green and Red Firefly Bioluminescence from 5,5-Dimethyloxyluciferin. *J. Am. Chem. Soc.* **2002**, *124*, 2112.
- [99] Hirano, T.; Hasumi, Y.; Ohtsuka, K.; Maki, S.; Niwa, H.; Yamaji, M.; Hashizume, D. Spectroscopic Studies of the Light-Color Modulation Mechanism of Firefly (Beetle) Bioluminescence. *J. Am. Chem. Soc.* **2009**, *131*, 2385.

- [100] van der Horst, M. A.; Hellingwerf, K. J. Photoreceptor Proteins, “Star Actors of Modern Times”: A Review of the Functional Dynamics in the Structure of Representative Members of Six Different Photoreceptor Families. *Acc. Chem. Res.* **2004**, *37*, 13.
- [101] Rockwell, N. C.; Lagarias, J. C. A Brief History of Phytochromes. *ChemPhysChem* **2010**, *11*, 1172.
- [102] Conrad, K. S.; Manahan, C. C.; Crane, B. R. Photochemistry of Flavoprotein Light Sensors. *Nat. Chem. Biol.* **2014**, *10*, 801.
- [103] Mathes, T.; Heilmann, M.; Pandit, A.; Zhu, J.; Ravensbergen, J.; Klotz, M.; Fu, Y.; Smith, B. O.; Christie, J. M.; Jenkins, G. I.; Kennis, J. T. M. Proton-Coupled Electron Transfer Constitutes the Photoactivation Mechanism of the Plant Photoreceptor UVR8. *J. Am. Chem. Soc.* **2015**, *137*, 8113.
- [104] Ziegler, T.; Möglich, A. Photoreceptor Engineering. *Front. Mol. Biosci.* **2015**, *2*, 30.
- [105] Möglich, A.; Moffat, K. Engineered Photoreceptors as Novel Optogenetic Tools. *Photochem. Photobiol. Sci.* **2010**, *9*, 1286.
- [106] Auldridge, M. E.; Satyshur, K. A.; Anstrom, D. M.; Forest, K. T. Structure-Guided Engineering Enhances a Phytochrome-Based Infrared Fluorescent Protein. *J. Biol. Chem.* **2012**, *287*, 7000.
- [107] Salomon, M.; Christie, J. M.; Knieb, E.; Lempert, U.; Briggs, W. R. Photochemical and Mutational Analysis of the FMN-Binding Domains of the Plant Blue Light Receptor, Phototropin. *Biochemistry* **2000**, *39*, 9401.
- [108] Shcherbakova, D. M.; Shemetov, A. A.; Kaberniuk, A. A.; Verkhusha, V. V. Natural Photoreceptors as a Source of Fluorescent Proteins, Biosensors, and Optogenetic Tools. *Annu. Rev. Biochem.* **2015**, *84*, 519.
- [109] Miesenböck, G. Optogenetic Control of Cells and Circuits. *Annu. Rev. Cell Dev. Biol.* **2011**, *27*, 731.
- [110] Toettcher, J. E.; Voigt, C. A.; Weiner, O. D.; Lim, W. A. The Promise of Optogenetics in Cell Biology: Interrogating Molecular Circuits in Space and Time. *Nat. Methods* **2011**, *8*, 35.
- [111] Huala, E.; Oeller, P. W.; Liscum, E.; Han, I.-S.; Larsen, E.; Briggs, W. R. Arabidopsis NPH1: A Protein Kinase with a Putative Redox-Sensing Domain. *Science* **1997**, *278*, 2120.
- [112] Christie, J. M.; Reymond, P.; Powell, G. K.; Bernasconi, P.; Raibekas, A. A.; Liscum, E.; Briggs, W. R. Arabidopsis NPH1: A Flavoprotein with the Properties of a Photoreceptor for Phototropism. *Science* **1998**, *282*, 1698.

- [113] Christie, J. M.; Nozue, M. S. K.; Wada, M.; Briggs, W. R. LOV (Light, Oxygen, or Voltage) Domains of the Blue-Light Photoreceptor Phototropin (NPH1): Binding Sites for the Chromophore Flavin Mononucleotide. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 8779.
- [114] Jarillo, J. A.; Gabrys, H.; Capel, J.; Alonso, J. M.; Ecker, J. R.; Cashmore, A. R. Phototropin-Related NPL1 Controls Chloroplast Relocation Induced by Blue Light. *Nature* **2001**, *410*, 952.
- [115] Kagawa, T.; Sakai, T.; Suetsugu, N.; Oikawa, K.; Ishiguro, S.; Kato, T.; Tabata, S.; Okada, K.; Wada, M. Arabidopsis NPL1: A Phototropin Homolog Controlling the Chloroplast High-Light Avoidance Response. *Science* **2001**, *291*, 2138.
- [116] Christie, J. M. Phototropin Blue-Light Receptors. *Annu. Rev. Plant Biol.* **2007**, *58*, 21.
- [117] Losi, A.; Gärtner, W. Old Chromophores, New Photoactivation Paradigms, Trendy Applications: Flavins in Blue Light-Sensing Photoreceptors. *Photochem. Photobiol.* **2011**, *87*, 491.
- [118] 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* **2008**, *198*, 221.
- [119] 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* **2009**, *113*, 15610.
- [120] Sikorska, E.; Khmelinskii, I. V.; Koput, J.; Sikorska, M. Electronic Structure of Lumiflavin and Its Analogues in their Ground and Excited States. *J. Mol. Struct. (THEOCHEM)* **2004**, *676*, 155.
- [121] Climent, T.; González-Luque, R.; Merchán, M.; Serrano-Andrés, L. Theoretical Insight into the Spectroscopy of Isoalloxazine, the Flavin Core Ring. *J. Phys. Chem. A* **2006**, *110*, 13584.
- [122] Wagner, J. R.; Zhang, J.; Brunzelle, J. S.; Vierstra, R. D.; Forest, K. T. High Resolution Structure of *Deinococcus* Bacteriophytochrome Yields New Insight into Phytochrome Architecture and Evolution. *J. Biol. Chem.* **2007**, *282*, 12298.
- [123] Essen, L.-O.; Mailliet, J.; Hughes, J. The Structure of a Complete Phytochrome Sensory Module in the Pr Ground State. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 14709.
- [124] Spillane, K. M.; Dasgupta, J.; Mathies, R. A. Conformational Homogeneity and Excited-State Isomerization Dynamics of the Bilin Chromophore in Phytochrome Cph1 from Resonance Raman Intensities. *Biophys. J.* **2012**, *102*, 709.

- [125] Song, C.; Psakis, G.; Kopycki, J.; Lang, C.; Matysik, J.; Hughes, J. The D-ring, Not the A-ring, Rotates in *Synechococcus* OS-B' Phytochrome. *J. Chem. Phys.* **2014**, *289*, 2552.
- [126] Altoè, P.; Climent, T.; Fusco, G. C. D.; Stenta, M.; Bottoni, A.; Serrano-Andrés, L.; Merchán, M.; Orlandi, G.; Garavelli, M. Deciphering Intrinsic Deactivation/Isomerization Routes in a Phytochrome Chromophore Model. *J. Phys. Chem. B* **2009**, *113*, 15067.
- [127] Durbeej, B. On the Primary Event of Phytochrome: Quantum Chemical Comparison of Photoreactions at C₄, C₁₀ and C₁₅. *Phys. Chem. Chem. Phys.* **2009**, *11*, 1354.
- [128] Strambi, A.; Durbeej, B. Initial Excited-State Relaxation of the Bilin Chromophores of Phytochromes: A Computational Study. *Photochem. Photobiol. Sci.* **2011**, *10*, 569.
- [129] Zhuang, X.; Wang, J.; Lan, Z. Tracking of the Molecular Motion in the Primary Event of Photoinduced Reactions of a Phytochromobilin Model. *J. Phys. Chem. B* **2013**, *117*, 15976.
- [130] Wagner, J. R.; Zhang, J.; von Stetten, D.; Günther, M.; Murgida, M.; Mro-ginski, A.; Walker, J. M.; Forest, K. T.; Hildebrandt, P.; Vierstra, R. D. Mutational Analysis of *Deinococcus radiodurans* Bacteriophytochrome Reveals Key Amino Acids Necessary for the Photochromicity and Proton Exchange Cycle of Phytochromes. *J. Biol. Chem.* **2008**, *283*, 12212.
- [131] Neugebauer, J.; Veldstra, J.; Buda, F. Theoretical Spectroscopy of Astaxanthin in Crustacyanin Proteins: Absorption, Circular Dichroism, and Nuclear Magnetic Resonance. *J. Phys. Chem. B* **2011**, *115*, 3216.

CHAPTER 6

Publications Included in the Thesis

- I On the Origin and Variation of Colors in Lobster Carapace**
S. Begum, M. Cianci, B. Durbeej, O. Falklöf, A. Hädener, J. R. Helliwell, M. Helliwell, A. C. Regan and C. I. F. Watt
Phys. Chem. Chem. Phys. **2015**, *17*, 16723
- II Distinguishing Between Keto-Enol and Acid-Base Forms of Firefly Oxyluciferin Through Calculation of Excited-State Equilibrium Constants**
O. Falklöf and B. Durbeej
J. Comput. Chem. **2014**, *35*, 2184
- III Inter-Excited State Phosphorescence in the Four-Component Relativistic Kohn-Sham Approximation: A Case Study on Lumiflavin**
O. Falklöf, B. Durbeej and P. Norman
Accepted for publication in J. Phys. Chem. A, DOI: 10.1021/acs.jpca.5b08908
- IV Red-Light Absorption and Fluorescence of Phytochrome Chromophores: A Comparative Theoretical Study**
O. Falklöf and B. Durbeej
Chem. Phys. **2013**, *425*, 19
- V Modeling of Phytochrome Absorption Spectra**
O. Falklöf and B. Durbeej
J. Comput. Chem. **2013**, *34*, 1363
- VI QM/MM Modeling Highlights the Importance of Steric Effects in the Photoactivation of a Bacteriophytochrome**
O. Falklöf and B. Durbeej
Manuscript

My Contributions to the Papers

I was responsible for carrying out all calculations for all papers. For Papers II and IV-VI, I planned the research, analyzed the results and wrote the manuscript together with Bo Durbeej. For Paper III, I planned the research, analyzed the results and wrote the manuscript together with Patrick Norman and Bo Durbeej. For Paper I, I planned and analyzed the theoretical work together with Bo Durbeej, and wrote the manuscript together with all co-authors.

Papers

The articles associated with this thesis have been removed for copyright reasons. For more details about these see:

<http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-122614>