Photochemical properties of phytochrome and firefly luciferase chromophores: A theoretical study

Olle Falklöf
Abstract

This licentiate thesis presents computational chemistry studies on photochemical properties of phytochrome and firefly luciferase chromophores.

Phytochromes are bilin-containing proteins that based on the ambient light environment regulate a number of physiological and developmental processes in bacteria, cyanobacteria, fungi and plants. From the viewpoint of computational modeling, however, only a few studies have been devoted to these systems. In this thesis, two systematic studies comparing calculated and experimental UV-vis spectra of bilin chromophores in protein and solution environments are presented. The first study focuses on how hybrid quantum mechanics/molecular mechanics methods are best applied to calculate absorption spectra of a bacteriophytochrome. The second study, in turn, investigates the performance of a number of quantum chemical methods in calculating absorption and emission spectra of sterically locked bilin chromophores.

Firefly luciferase catalyzes a chemical reaction in which the electronically excited oxyluciferin is formed and subsequently emits light. Depending on the conditions, oxyluciferin can exist in a number of different chemical forms. To date, there is no consensus regarding which of these that most significantly contributes to the light emission. In this thesis, the most probable form of the light emitter is predicted by calculating excited-state $pK_E$ and $pK_A$ values, in aqueous solution, of the various equilibrium reactions relevant for the oxyluciferin system.
First of all, I would like to thank my supervisor Bo Durbeej for his excellent supervision and guidance. I would also like to thank Patrick Norman for being my co-supervisor and for future collaboration!

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

I am also thankful to all present and former colleagues in the Computational Physics group for making my working days more pleasant. My special thanks go to Chang-Feng Fang and Baswanth Oruganti.

Ett stort tack går till mamma, pappa, Anneli och Anders för ert stöd och för all hjälp under skrivandet av den här avhandlingen. Avslutningsvis vill jag tacka de som står mig allra närmast: Joanna och Julie, ni är ljuset i mitt liv!

Olle Falklöf
Linköping, January 2014
# Contents

1 Introduction 1

2 Quantum Chemistry 3
   2.1 Basic Approximations .......................... 3
   2.2 Hartree-Fock Theory ............................ 5
   2.3 Density Functional Theory ..................... 6
      2.3.1 The Kohn-Sham Equations .................. 7
      2.3.2 Exchange-Correlation Functionals ........... 9

3 Calculation of Excited States 11
   3.1 Time-Dependent Density Functional Theory .......... 11
   3.2 Other Methods .................................. 14

4 Multiscale Models 17
   4.1 QM/MM Methods ................................ 17
      4.1.1 The ONIOM Approach ......................... 18
   4.2 Polarizable Continuum Models ................... 21
      4.2.1 Non-Equilibrium Solvation in TD-DFT ...... 22

5 Phytochrome 25
   5.1 Paper I ........................................ 26
   5.2 Paper II ....................................... 30

6 Firefly luciferase 33
   6.1 Paper III ...................................... 34

7 Conclusions 37

Bibliography 39
<table>
<thead>
<tr>
<th>8 Publications</th>
<th>43</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paper I</td>
<td>45</td>
</tr>
<tr>
<td>Paper II</td>
<td>59</td>
</tr>
<tr>
<td>Paper III</td>
<td>71</td>
</tr>
</tbody>
</table>
Light is an essential part of life on Earth. In Nature, there are many examples of systems that are governed by light such as the chloroplasts in plants converting carbon dioxide into carbohydrates and the human eyes sending electrical impulses to the brain. There are also many examples of organisms that produce light through chemical reactions inside their bodies such as colorful jellyfish in the oceans and the glowing fireflies lighting up the night sky. These examples illustrate some of the phenomena that are studied in theoretical photochemistry.

Theoretical photochemistry is a relatively young field of science. In the early 1950’s, Rudolph Pariser, Robert Parr and John Pople developed one of the first methods to calculate and predict UV-vis spectra of unsaturated molecules. [1, 2] Although their method offered good performance for $\pi \rightarrow \pi^*$ excitations within the valence shell, it was not clear how to treat other types of transitions. For a more general treatment of excited states, new theoretical approaches were needed.

In the 1980’s, multiconfigurational self-consistent field methods such as the complete active-space self-consistent field [3] approach became available for calculating excited states. With this method one was able to explore excited-state potential energy surfaces utilizing analytical gradient techniques that besides calculation of spectra also enabled studies on photochemical reactions. While the complete active-space self-consistent field approach in many cases gave good results, the method was limited in applicability to rather small systems. For modeling large molecular systems, on the other hand, configuration interaction singles approach evolved to the standard choice of method. [4] Unfortunately, the performance of this method was actually rather poor.

A breakthrough in the modeling of excited states of large molecules was the advent of time-dependent density functional theory, [5–9] which had similar computational requirements as the configuration interaction singles approach but performed markedly better. Furthermore, just like its parent theory (density func-
Today, TD-DFT has become the standard method of choice for calculation of excited states of large molecular systems. However, despite significant methodological improvements during the past decades, it is still a very challenging task to describe photochemical phenomena. This fact, together with all the fascinating applications, makes theoretical photochemistry an exciting field of research.

The present licentiate thesis deals with theoretical modeling of photochemical properties of biological chromophores. In particular, quantum mechanical methods are applied to study the absorption and emission properties of the bilin chromophore in phytochromes, and to investigate the excited-state equilibria of the oxyluciferin light-emitter in firefly luciferase. The thesis is divided into several chapters. In Chapters 2 – 4, the theory underlying the methods used are introduced. Chapter 5 and 6 present the background to the different projects and summarize the key results of the respective papers. Finally, conclusions are given in Chapter 7.
Quantum chemistry is a branch of chemistry in which chemical phenomena are studied based on the fundamental laws of quantum mechanics. During the past century, a plethora of theoretical methods have been developed with the aim of explaining and predicting chemical phenomena. One class of methods, the so-called \textit{ab initio} methods, are based on the electronic wave function of the system of interest. Other methods are instead based on the electron density. In this chapter, some of the basic approximations underlying the wave-function and electron-density based methods used in this thesis will be introduced.

### 2.1 Basic Approximations

The starting point for a quantum mechanical description of a molecular system consisting of $M$ nuclei and $N$ electrons is the time-dependent Schrödinger equation

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

where $\Psi$ is the wave function of the system. The Hamiltonian operator, $\hat{H}$, is given by

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

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 interelectronic, internuclear and electron-nucleus interaction operators, respectively. If the Hamiltonian is time-independent, a solution of the time-dependent Schrödinger equation takes the form

$$\Psi(R, r, t) = \psi(R, r) e^{-iEt/\hbar}. \quad (2.3)$$
Here, the wave function $\psi(R, r)$ describes a stationary state with energy $E$, and $r$ and $R$ 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,

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

(2.4)

can be derived. This is one of the most central equations in quantum chemistry.

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

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

(2.5)

In this expression, nuclear wave functions $\psi_{ni}(R)$ are used as expansion coefficients and $\psi_{ei}(r; R)$ depends explicitly on the electronic coordinates and parametrically on the nuclear positions. Multiplying Eq. 2.5 to the left with the electronic wave function of state $j$ and integrating over the electronic coordinates results in

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

(2.6)

where the electronic energy ($E_{ej} = E_{ej}(R)$) is a function of the nuclear geometry and $\Lambda$ represents the 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

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

(2.7)

for the total energy $E$ can be obtained.

To further simplify the calculations, the Born-Oppenheimer (BO) approximation states that, due to the large mass difference between an electron and each nucleus, the $\langle \psi_{ej}|\hat{T}_n|\psi_{ej}\rangle$ term can be neglected if the electronic wave function is a slowly varying function of the nuclear coordinates. Both the adiabatic and BO approximations are usually good approximations. However, for nearly degenerate states such as close to conical intersections these approximations break down. The resulting equation after the adiabatic and BO approximations have been introduced is

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

(2.8)

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

$$\hat{H}_e\psi_e(r; R) = E_{ej}(R)\psi_e(r; R).$$

(2.9)

Another consequence of the large mass difference 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. Furthermore, the total internuclear repulsion
2.2 Hartree-Fock Theory

does not depend on the electronic coordinates. Therefore, the nuclear repulsion energy can be added once the electronic problem has been solved. The electronic Hamiltonian can thereby be written as

\[ H_e = T_e + V_{en} + V_{ee}. \]

To solve the Schrödinger equation in practical calculations, the computation can be separated into two steps. First, the electronic Schrödinger equation is solved for a fixed nuclear geometry. Then, the resulting energy, \( E_{ej} \), is used for solving the nuclear problem. In this thesis, the focus is on solving the electronic Schrödinger equation (Eq. 2.9). To simplify the notation, the electronic Hamiltonian, wave functions and energies are hereafter denoted as

\[ \hat{H}, \psi \] and \( E \), respectively, and Eq. 2.9 thus reads

\[ \hat{H}\psi = E\psi. \ 
\]

Unfortunately, analytic solutions of Eq. 2.10 can only be derived for very simple (i.e., one-electron) systems. For other systems that include electron-electron interactions, further approximations are required.

2.2 Hartree-Fock Theory

In the Hartree-Fock (HF) approximation, the electron-electron interactions are treated in an average fashion. This means that every single electron interacts with a mean 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,

\[ \psi_{HF}(x_1, x_2, x_3, \ldots) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_1(x_1) & \phi_1(x_2) & \ldots & \phi_1(x_N) \\ \phi_2(x_1) & \phi_2(x_2) & \ldots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \phi_N(x_1) & \ldots & \ldots & \phi_N(x_N) \end{vmatrix}, \]

the HF equations are obtained as

\[ \hat{f}(i)\phi_i(x_i) = \epsilon_i\phi_i(x_i), \]

by applying the variational principle. Here, \( \epsilon_i \) is the energy of the \( i \)th state described by the \( \phi_i(x_i) \) spin orbital. The 4-dimensional coordinate \( x_i \) represents both spatial (\( r_i \)) and spin (\( \sigma_i \)) variables. The Fock operator,

\[ \hat{f}(i) = -\frac{1}{2}\nabla_i^2 + \hat{v}_{HF}(i), \]

is an effective one-electron operator that includes an effective one-electron potential operator \( \hat{v}_{HF}(i) = \hat{v}_{en}(i) + \hat{J}(i) + \hat{K}(i) \). \( \hat{v}_{HF}(i) \) describes the interactions between an electron with all nuclei and all other electrons. While the Hartree (Coulomb) operator, \( \hat{J}(i) \), is a local operator and corresponds to the classical electrostatic repulsion, the exchange operator, \( \hat{K}(i) \), 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).

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

$$E_{\text{HF}} = \sum_{i=1}^{N} \epsilon_i - \frac{1}{2} V_{\text{ee}} + V_{\text{nn}} = \sum_{i=1}^{N} \epsilon_i - \frac{1}{2} (J - K) + V_{\text{nn}}. \quad (2.14)$$

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 orbital energies.

While HF often capture a large part of the total energy (with a large basis set), \cite{10} the remaining energy corresponding to correlated electronic motions is often important to describe chemical reactions. In a given basis set, a standard way of defining the correlation energy ($E_{\text{corr}}$) is

$$E_{\text{corr}} = E_0 - E_{\text{HF}}, \quad (2.15)$$

where $E_0$ is the exact non-relativistic electronic energy and $E_{\text{HF}}$ is the HF energy. 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 helium atom. Examples of ab initio methods that have been developed to treat dynamical correlation effects are configuration interaction (CI), coupled cluster (CC) and many-body perturbation theory (MBPT) 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, multideterminantal methods are required such as the complete active-space self-consistent field (CASSCF) method. \cite{3} If both dynamical and non-dynamical are important, the dynamical correlation energy can be added on top of a CASSCF energy by using second-order perturbation theory (CASPT2). \cite{11}

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

### 2.3 Density Functional Theory

Density functional theory (DFT) offers a possible route for carrying out correlated calculations at a relatively low computational cost. The central quantity in DFT is the electron density ($\rho(r_1)$), which is related to the many-electron wave function through

$$\rho(r_1) = N \int |\psi(x_1, x_2, ..., x_n)|^2 dx_1 dx_2 ... dx_n. \quad (2.16)$$
The many-electron Hamiltonian (in atomic units) for a system consisting of \( \mathcal{N} \) electrons is explicitly given by

\[
\hat{H} = -\sum_{i=1}^{\mathcal{N}} \frac{1}{2} \nabla^2_i + \sum_{i=1}^{\mathcal{N}} v_{\text{ext}}(r_i) + \sum_{i=1}^{\mathcal{N}} \frac{1}{r_{ij}},
\]  

(2.17)

where the number of electrons, \( \mathcal{N} \), follows straightforwardly from the electron density as

\[
\mathcal{N} = \int \rho(r)dr.
\]  

(2.18)

In a pioneering work by Hohenberg and Kohn, [12] the foundation of DFT was established. In this work, they proved that the external potential \( v_{\text{ext}}(r) \) in Eq. 2.17 is uniquely defined (up to an additive constant) by the electron density. This is the so-called first Hohenberg-Kohn (HK) theorem. [12] The consequence of this theorem is that the Hamiltonian can be fully recreated from the electron density \( \hat{H} = \hat{H}[\rho(r)] \) and that the electron density determines the wave function and all properties of the system at hand. In particular, the exact electronic energy of a system can be written as

\[
E[\rho] = T[\rho] + V_{\text{ee}}[\rho] + V_{\text{ext}}[\rho] = F[\rho] + \int \rho(r)v_{\text{ext}}(r)dr,
\]  

(2.19)

where \( F[\rho] \) is a (unknown) universal functional of \( \rho \)

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

(2.20)

For a given external potential \( v_{\text{ext}}(r) \), they also showed the exact ground-state electron density \( \rho_0 \), derived from an \( \mathcal{N} \)-electron antisymmetric wave function, can be obtained by variationally minimizing the energy, \( E_0 = E[\rho_0] \). This is the so-called second HK theorem. [12]

### 2.3.1 The Kohn-Sham Equations

Thomas-Fermi (TF) theory is considered to be the first density functional. In this theory, the kinetic energy is given by a local functional of the electron density. While the TF functional reasonably reproduce atomic energies, TF theory is unable to describe important chemical phenomena such as atomic shell structure and covalent bonds. [13] The main flaw in TF theory is the way electron dynamics is incorporated into the functional, which cause large errors in the kinetic energy. One way to improve the TF model is to include (non-local) kinetic energy terms such as the von Weizäcker correction. [14] Although the inclusion of non-local terms improves the prediction 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 \( \rho \), Kohn and Sham [15] introduced a practical scheme to calculate the electronic energy using one-electron wave functions. In this scheme, Kohn and Sham proposed that a (real) system of interacting electrons moving in an external potential
Quantum Chemistry can be replaced by a fictitious system of non-interacting electrons moving in an effective potential 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 reads

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

(2.21)

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

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

(2.22)

The key assumption in Kohn-Sham theory is to divide the complicated (real) kinetic energy functional \( T[\rho] \) 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 Kohn-Sham scheme is thus defined as

\[ F[\rho] = T_S[\rho] + J[\rho] + E_{\text{xc}}[\rho], \]  

(2.23)

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

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

(2.24)

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

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

(2.25)

The third term in Eq. 2.23 \( (E_{\text{xc}}) \) is the exchange-correlation energy of the real system, which is defined as

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

(2.26)

Minimizing the exact energy of the real (interacting) system

\[ E[\rho] = \int \rho(\mathbf{r}) v_{\text{ext}}(\mathbf{r}) d\mathbf{r} + T_S[\rho] + J[\rho] + E_{\text{xc}}[\rho], \]  

(2.27)

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

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

(2.28)
2.3 Density Functional Theory

Solving the Kohn-Sham equations using this potential gives the exact electronic energy of the system within the BO approximation. Unfortunately, since the exact analytic form of the exchange-correlation potential ($E_{xc}[\rho]$) is unknown for an arbitrary density, this potential 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 compared to orbital-free DFT such as Thomas-Fermi theory. In the next section, some general ideas of these approximations will be discussed.

2.3.2 Exchange-Correlation Functionals

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

The functionals can be divided into several classes. 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 into global and long-range-corrected hybrid functionals. While a global hybrid functional contains a fixed amount of HF exchange, the fraction of HF exchange included in a long-range-corrected hybrid functional depends on the interelectronic distance. Since only the long-range-corrected hybrid functionals exhibit the correct asymptotic distance dependence ($-1/r$), they usually give much better excitation energies of Rydberg and charge-transfer states than the other functionals.
CHAPTER 3

Calculation of Excited States

The DFT formalism based on the HK theorems [12] can in principle be generalized to include the lowest state of a given space-spin symmetry, whereby one perform separate calculations on the ground and excited states to obtain excitation energies (the so-called ∆SCF approach). Although the ∆SCF approach can be useful if the excited state of interest is of a different symmetry than that of the ground state, this methodology breaks down for asymmetric molecules. For such systems, more general theoretical methods are necessary.

One such method is time-dependent density functional theory (TD-DFT), [5–9] which has become a very popular method for modeling spectroscopic and photochemical properties of large molecules. TD-DFT perform very well in many applications, however, this method has also some well-documented disadvantages. In this chapter, the basic ideas of TD-DFT and some other alternative methods will be presented.

To distinguish various types of orbitals used, indices $i$ and $j$ refer to occupied orbitals, indices $a$ and $b$ refer to virtual orbitals, and $k$, $l$, $m$ and $n$ refer to general orbitals throughout this chapter.

3.1 Time-Dependent Density Functional Theory

In 1984, Runge and Gross [5] extended the basic ideas from static DFT to 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, $\rho(r, t)$, determines the external potential up to an additive function of time. They also derived three practical schemes to calculate this density.

For an interacting system of electrons subject to a time-dependent external
the so-called linear response function \( \chi \) response DFT. The adiabatic approximation seems to work best for low-lying excited states. [8] makes it possible to use the standard exchange-correlation functionals of static \( A \) is introduced. In this approximation, \( v \) is the time-dependent exchange-correlation functional, which is the analogue of \( E_{\text{xc}} \) of DFT. Normally, the adiabatic approximation

\[
v_{\text{xc}}[\rho](r, t) = \frac{\delta E_{\text{xc}}[\rho]}{\delta \rho(r)} = v_{\text{xc}}[\rho(t)](r) \tag{3.4}
\]

is introduced. In this approximation, \( A_{\text{xc}} \) is assumed to be local in time, which makes it possible to use the standard exchange-correlation functionals of static DFT. The adiabatic approximation seems to work best for low-lying excited states. [8]

If the perturbation \( v_1(r, t) \) 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 \( \chi_S(t, t', r, r') \) of the noninteracting system and takes the form

\[
\rho^{(1)}(r, t) = \int \chi_S(t, t', r, r') \left[ v_1(r', t') + \int \frac{\rho^{(1)}(r'', t')}{|r' - r''|} dr'' \right] dr' dt',
\tag{3.5}
\]

where \( \chi_S(t, t', r, r') \) can explicitly be expressed in terms of the Kohn-Sham orbitals as

\[
\chi_S(\omega, r, r') = \sum_{i,a} \left[ \frac{\phi_i^*(r) \phi_a(r') \phi_i^*(r') - \phi_i(r) \phi_a^*(r') \phi_a^*(r)}{\omega - (\epsilon_a - \epsilon_i)} \right].
\tag{3.6}
\]

Since \( \rho^{(1)} \) depends of \( \rho^{(1)} \) itself, Eq. 3.5 has to be solved self-consistently with an iterative scheme.

Taking the Fourier transform with respect to time, one can transform Eq. 3.5 to the exact frequency-dependent linear density response within the adiabatic approximation,

\[
\rho^{(1)}(r, \omega) = \int \chi_{KS}(\omega, r, r') \left[ v_1(r', \omega) + \int \frac{\rho^{(1)}(r'', \omega)}{|r' - r''|} dr'' \right] dr' + \int f_{\text{xc}}(r', r'') \rho^{(1)}(r'', \omega) dr''.
\tag{3.7}
\]
3.1 Time-Dependent Density Functional Theory

where the exchange-correlation kernel,

\[ f_{xc}(r', r'') = \frac{\delta^2 E_{xc}}{\delta \rho(r') \delta \rho(r'')} , \tag{3.8} \]

has been introduced.

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, \( \omega = \epsilon_a - \epsilon_i \), between Kohn-Sham orbital energies for which \( \chi_{KS}(\omega, r, r') \) has poles. [7]

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

\[ \rho^{(1)}(r, \omega) = \sum_{i,a} [X_{ia}(\omega)\phi_a^*(r)\phi_i(r) + Y_{ia}(\omega)\phi_a(r)\phi_i^*(r)] , \tag{3.9} \]

with the ground-state Kohn-Sham orbitals. Using this parametrization together with the expression for \( \chi_{KS}(\omega, r, r') \) above (Eq. 3.6), the generalized eigenvalue problem

\[ \begin{pmatrix} A & B \\ B^* & A^* \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix} = \omega \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix} \tag{3.10} \]

can be derived, [6] where the eigenvalue \( \omega = \omega_0 = (E_I - E_0)/\hbar \) corresponds to the true vertical excitation energy from the ground state to the excite state \( I \). The matrix elements of \( A \) and \( B \), in turn, are defined as

\[ A_{ia,jb} = \delta_{ab} \delta_{ij} (\epsilon_a - \epsilon_i) + K_{ia,jb} \tag{3.11} \]

and

\[ B_{ia,jb} = K_{ia,bj} \tag{3.12} \]

where

\[ K_{kl,mn} = \int \phi_k^*(r) \phi_l(r) \left( \frac{1}{|r - r'|} + f_{xc}(r', r'') \right) \phi_m^*(r') \phi_n(r') \, dr' \, dr. \tag{3.13} \]

The exact structures of \( A \) and \( B \) depend on the specific exchange-correlation functional used.

For real orbitals (typically used in practical calculations), the matrix \( (A - B) \) becomes positive definite and the dimensionality of the eigenvalue problem can be reduced by reformulating Eq. 3.10 as

\[ (A - B)^{1/2} (A + B) (A - B)^{1/2} (X + Y)' = \omega^2 (X + Y)' , \tag{3.14} \]

where \( (X + Y)' = (A - B)^{1/2} (X + Y) \). [9]

Eq. 3.14 (or generally Eq. 3.10) provides a route for obtaining all excitation energies in one single calculation without explicitly considering the excited-state.
Calculation of Excited States

The transition dipole moment between the ground state and excited state $I$, $\langle \psi_0 | \hat{\mu} | \psi_I \rangle$, can be used to construct the oscillator strengths,

$$f_{10} = \frac{2}{3} \omega_{10} |\langle \psi_0 | \hat{\mu} | \psi_I \rangle|^2,$$

(3.15)
can also be obtained from the ground-state wave function. [8] Furthermore, it is also possible to derive analytical gradients without explicitly considering the excited-state wave functions. [16–19] Consequently, all ingredients for studying photochemical and spectroscopic properties can be obtained from the ground-state wave function.

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

Today, TD-DFT has become the primary tool for calculation of 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 long-range-corrected hybrid functionals such as LC-BLYP, [20] CAM-B3LYP, [21] and ωB97X-D, [22] it is also of interest to consider some alternative methods available for studying excited states of large molecules.

3.2 Other Methods

The simplest ab initio method for treating excited states is the configuration interaction singles (CIS) approach. [4] Starting from the optimized HF wave function, the CIS wave function is constructed as a linear combination of all singly excited determinants, which have been obtained by replacing one occupied orbital by one virtual orbital. In applications, CIS, which typically overestimate excitation energies by more than 1 eV, [23] is usually inferior to TD-DFT. The large errors in the CIS energies arise from the neglect of correlation effects together with the fact that the HF wave function 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 wave-function-based approaches that include correlation effects of the excited states are today available for modeling large systems. Examples of such methods are CIS with perturbation-theory doubles corrections (CIS(D)), [24] the second-order approximate coupled-cluster singles and doubles (CC2) [25] and the symmetry-adapted cluster configuration interaction (SAC-CI) methods. [26] CC2 is an approximation of the equation-of-motion coupled-cluster singles and doubles (EOM-CCSD) method, [27] whereas SAC-CI is similar to EOM-CCSD. Although the family of CC methods can systematically be improved by including higher excitation operators, SAC-CI is often the most sophisticated CC method available for calculations on large molecules.
3.2 Other Methods

If near degeneracies come into play, a multireference method such as the complete active space second-order perturbation theory (CASPT2) [11] is required. For such states both dynamical and non-dynamical correlation become important. While CASPT2 is regarded as the gold standard for small molecules, it is more challenging to obtain accurate results for large molecules since small basis sets and incomplete active spaces must be used. Therefore, if non-dynamical correlation effects are of minor importance, SAC-CI often provides a much better alternative.
One of the main issues in computational chemistry is to find a balance between the accuracy of the results and the computational cost. While quantum chemical methods can give a detailed description of the electronic structure, they are often restricted to treatment of systems consisting of up to a few hundreds of atoms. Classical mechanics, on the other hand, cannot describe chemical reactions but can be used to simulate large systems such as proteins. The basic idea in multiscale modeling is to divide a system of interest into a number of subsystems treated at different levels of theory.

In this chapter, two multiscale models will be introduced. The first, termed the hybrid quantum mechanics/molecular mechanics (QM/MM) method, combines a quantum mechanical treatment of the chemically relevant part of the system with a discrete classical description of the remaining part of the system. The second, termed polarizable continuum model, replaces the discrete nature of the surrounding environment by a dielectric continuum.

4.1 QM/MM Methods

The hybrid quantum mechanics/molecular mechanics (QM/MM) approach has become a valuable tool in modeling complex biomolecular systems such as photosensory proteins, enzymes and DNA. In the QM/MM approach, the system ($S$) of interest is divided into an inner ($I$) subsystem treated at a quantum mechanical (QM) level of theory and an outer ($O$) subsystem modeled with a molecular mechanics (MM) force field. This is illustrated in Fig. 4.1.

One of the most difficult tasks in setting up a QM/MM calculation is the selection of $I$ 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
Figure 4.1. Schematic partitioning of the entire system (S) into inner (I) and outer (O) subsystems. The blue ring represents the boundary region.

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.

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

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

(4.1)

or a subtractive

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

(4.2)

scheme. Here, L and I+L represent link atoms and the capped QM system, respectively.

4.1.1 The ONIOM Approach

The ONIOM (Our N-layered Integrated molecular Orbital and molecular Mechanics) approach [28] is based on the subtractive scheme and is an example of a method that can be used to carry out QM/MM calculations. 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. In the original formulation of
4.1 QM/MM Methods

ONIOM(QM:MM), the $E_{\text{sub}}^{\text{QM/MM}}(S)$ energy is obtained from the $E_{\text{QM}}(I+L)$ energy computed with a QM method, and the $E_{\text{MM}}(O)$ and $E_{\text{MM}}(I+L)$ energies calculated with a MM force field. Adopting the notation by Cornell et. al. [29] a typical force field energy can be expressed as

$$E_{\text{MM}} = E_r + E_{\theta} + E_{\phi} + E_{\text{vdW}} + E_{\text{el}},$$

where the bond stretching term is given by

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

the bond bending term by

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

and the dihedral angle term by

$$E_{\phi} = \sum_{\text{dihedrals}} V_n \left(1 + \cos(n\phi - \gamma)\right).$$

While these terms involve covalent bonds, the final two terms

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

and

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

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.

The electrostatic embedding (EE) scheme provides a more accurate way of treating the electrostatic QM-MM coupling term. 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

$$\hat{H}_{\text{EE}} = -\sum_{i=1}^{n} \sum_{J=1}^{M} \frac{s_J q_J}{|R_J - r_i|} + \sum_{I=1}^{N} \sum_{J=1}^{M} \frac{Z_I s_J q_J}{|R_J - R_I|},$$

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 $R_J$, $Z_I$ is the nuclear charge of QM nucleus $I$, and $R_I$ and $r_i$ represent the nuclear and electronic coordinates,
respectively. The indices \(i\), \(I\) and \(J\) run over the \(n\) QM electrons, the \(N\) 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 \(E_{\text{MM}}(I+L)\) energy, which thus is given by the expression

\[
E_{\text{EE}}^{\text{MM}}(I+L) = E_{\text{MM}}(I+L) + \sum_{I=1}^{N} \sum_{J=1}^{M} \frac{q_I s_J q_J}{|R_J - R_I|}. \tag{4.10}
\]

Here, the scaling factor \(s_J\) is identical to that in Eq. 4.9. Including electrostatic embedding, the QM/MM energy finally reads

\[
E_{\text{sub,EE}}^{\text{QM/MM}}(S) = E_{\text{MM}}(O) + E_{\text{EE}}^{\text{MM}}(I+L) - E_{\text{EE}}^{\text{MM}}(I+L). \tag{4.11}
\]

While electrostatic embedding offers a substantial improvement compared with the classical mechanical embedding scheme, the general issue concerning the compatibility between the MM charge 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 quantum mechanical method responds to point charges. For example, it has been shown for rhodopsins that density functional methods are not able to fully account for the electrostatic effects. [30]

Despite these concerns, the QM/MM methodology has become the state-of-the-art computational technique for modeling complex biomolecular systems.

While \(O\) is described by an atomistic model within the QM/MM formalism, it is sometimes possible to replace this discrete description by a polarizable dielectric continuum model. In the next section, the basic ideas of these so-called polarizable continuum models will be presented.
4.2 Polarizable Continuum Models

In a polarizable continuum model (PCM), a solvated molecule is described at a QM level of theory inside an electrostatic cavity surrounded by a polarizable dielectric continuum (characterized by a dielectric constant \( \epsilon \)) that represents the solvent (as illustrated in Fig. 4.2). In PCM, the solute induces a charge distribution on the cavity surface, which in turn interacts with the solute and polarizes the 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.

![Figure 4.2. Schematic illustration of the PCM model.](image)

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 \( s_i \). Based on these variables, the electronic perturbation operator can be expressed as

\[
\hat{V}_\sigma = \sum_{\text{tesserae}} \frac{q_i}{|s_i - r|},
\]

(4.12)

where \( r \) represents the electronic coordinates.

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

\[
G = E[\rho^0] + \Delta G_{\text{solv}},
\]

(4.13)

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

\[
\Delta G_{\text{solv}} = G_{\text{cav}} + G_{\text{disp-rep}} + G_{\text{el}} - E^0
\]

(4.14)

is the free energy of solvation. In this expression, \( G_{\text{cav}} \) 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 represent the non-electrostatic part of the free energy.

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

$$G_{el} = E[\rho] + G_{int},$$

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

$$G_{int} = \frac{1}{2} \sum_i q_i V_i$$

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 $\rho$, $q_i$, and $V_i$ can be determined by variationally minimizing of

$$G_{el} = \langle \psi | \hat{H}^0 + \frac{1}{2} V_\sigma | \psi \rangle$$

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

### 4.2.1 Non-Equilibrium Solvation in TD-DFT

The dynamic nature of the solvent plays an important role in many 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

$$q_{i,f} = \frac{\epsilon_f - 1}{\epsilon - 1} q_i$$

and a slow component

$$q_{i,s} = \frac{\epsilon - \epsilon_f}{\epsilon - 1} q_i,$$

where $\epsilon_i$ is a 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.

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

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

where $\hat{H}_{KS}^0$ is the time-dependent Kohn-Sham operator in the gas phase. While the $v^{PCM}$ potential depends on $\epsilon$ if both fast and slow degrees of freedom have time
to relax (equilibrium solvation), $\epsilon_f$ is used instead of $\epsilon$ if only the fast component is equilibrated (non-equilibrium solvation). [31]

In the SS implementation of PCM/TD-DFT, the electrostatic interaction part of the free energy of the excited state (2) explicitly depends on the ground-state (1) density and is given by the expression

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

$$+ \frac{1}{2} \left( \sum_i q_{i,s}^{(1)} V_{i,f}^{(2)} - \sum_i q_{i,s}^{(1)} V_{i,s}^{(1)} \right).$$

(4.21)

In this expression, 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. Here, $V_f$ and $V_s$ correspond to the potentials relative to the fast and slow degrees of freedom 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. [32]
Phytochromes constitute a superfamily of photosensory proteins that based on the ambient light environment regulate a number of physiological and developmental processes such as seed germination, flowering time and shade avoidance in plants, and phototaxis in bacteria. The red/far-red absorbing linear tetrapyrrole (bilin) chromophore is crucial for the biological function of phytochromes and is an example of a conserved structural motif within this family. Apart from the characteristic strong absorption of red light (the so-called Q band), UV-vis spectra of bilin chromophores also exhibit a band in the blue region (the so-called Soret band). In vivo, the bilin chromophores are covalently bound through a thioether linkage to a cysteine residue of the apoprotein.

Figure 5.1. Structure of the biliverdin IXα chromophore. Reprinted from Paper I with permission from John Wiley & Sons. Copyright 2013 Wiley Periodicals, Inc.
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 predominates in the dark-adapted inactive state, whereas the biologically active state contains higher concentrations of the Pfr form. Absorption of red light by the Pr form triggers a phototransformation to the Pfr form. This process starts with an excited-state isomerization involving rotation of the D pyrrole ring of the bilin chromophore (Fig. 5.1), and proceeds via a number of metastable intermediates. Despite much progress in recent years, the underlying mechanisms of the Pr → Pfr conversion are still not fully understood.

Despite the important role phytochromes play in Nature, only a few computational studies have been reported. To assess the performance of different density functionals and QM/MM methods in modeling phytochromes, we performed two systematic studies comparing calculated and experimental spectroscopic data. In Paper I, we investigated how the choice of QM/MM methodology effects the resulting UV-vis spectra of a *Deinococcus radiodurans* bacteriophytochrome, which harbors a biliverdin IXα (BV) chromophore (Fig. 5.1). In Paper II, the performance of primarily TD-DFT but also some *ab initio* methods was assessed by computing absorption and emission maxima of four synthetic derivatives of the bilin chromophores. In the next two sections, the key findings from these papers will be briefly discussed. For a complete compilation of the results, the reader is referred to Papers I and II attached as appendices.

### 5.1 Paper I

To study how the choice of density functional for the QM system affects the computed Q and Soret-band absorption maxima of phytochromes, UV-vis spectra of a *Deinococcus radiodurans* bacteriophytochrome (*Dr*BphP) in the Pr form (Fig. 5.2) were obtained from 50 – 100 calculated excitation energies. In these calculations, the chromophore binding pocket (that include the BV chromophore and its neighboring residues and water molecules) had been relaxed (structure denoted BV_{Full+Site} in Paper I). The QM system (denoted QM1 in Paper I) consisting of the BV chromophore and the side chain of Cys24 was treated with TD-DFT in combination with the 6-31+G(d,p) basis set, whereas the MM part was modeled with the AMBER force field. The computed Q and Soret-band absorption maxima are summarized in Table 5.1.

Focusing initially on the Q band, we first note that the pure GGAs (BP86, BLYP, B97D and τ-HCTH) exhibit better agreement (< 0.15 eV) with the experimental Q-band maximum than the corresponding values for the hybrid GGAs (B3LYP and PBE0). While this observation contrasts with the average performances of pure and hybrid GGAs documented in an extensive benchmark study, [34] the same trend was actually observed in a QM study on the chemically related porphin molecule and also for the sterically locked BV chromophores studied in Paper II. [6] It is further interesting to note that while the pure GGAs deviate from the experimental value by less than 0.15 eV, the hybrid GGAs (B3LYP and PBE0) and also those that, in part, have been designed to treat excited states...
Continuing with the Soret band, we observe that the pure and hybrid GGAs more accurately reproduce the experimental value than the functionals that have been designed to treat excited states. Indeed, while the pure GGAs (BP86, BLYP, B97D, \(\tau\)-HCTH) and the hybrid GGAs (B3LYP and PBE0) closely match (< 0.13 eV) the experimental value, M06-HF, LC-BLYP and LC-\(\omega\)PBE blue-shift this peak by 0.5 – 0.6 eV (\(\sim\) 0.3 eV for CAM-B3LYP).

Having carried out QM/MM calculations on the DrBphP protein with a number of density functionals, it is also of interest to investigate how well these functionals respond to the electric field induced by charges of the protein environment. Therefore, 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. In the absence of an experimental Q-band absorption maximum in the gas phase, the reference value for the calculation was instead estimated as 1.65 eV using high-level \textit{ab initio} methods. Based on the estimated value in the gas phase (1.65 eV) and the experimental DrBphP value (1.77 eV), it is predicted that the protein environment...
blue-shifts the intrinsic gas-phase absorption by 0.12 eV ($\lambda_{\text{max}}^{(2)} \rightarrow \lambda_{\text{max}}^{(3)}$). To support this shift, supplementary QM and QM/MM calculations were performed using CIS(D)/6-31G(d,p). The computed values are given in Table 5.2.

Interestingly, despite the fact the pure GGAs exhibit better agreement with the experimental DrBphP Q-band peak, it is actually the $\lambda_{\text{max}}^{(2)} \rightarrow \lambda_{\text{max}}^{(3)}$ values calculated with the hybrid functionals that more closely match the corresponding reference value (0.12 eV). Specifically, while the hybrid functionals predict blue-shifts of 0.08 – 0.11 eV and also accurately reproduce the CIS(D) value (0.14 eV), the pure GGAs actually give small red-shifts of 0.03 – 0.04 eV. In this regard, a finding from a computational study on rhodopsin proteins demonstrating that in particular pure GGAs, but also to some extent hybrid GGAs, are not able to fully account for absorption shifts induced by atomic point charges could explain the differences between the functionals observed here. [30]

Up until this point, the focus has been on spectral tuning by the electrostatic environment. However, the protein also perturbs the geometry of the chromophore. To quantify the geometrical effects on the absorption maximum, the reference value for the calculation was also in this case estimated (1.55 eV) using high-level ab initio methods. Based on the estimated reference values, it was predicted that protein-induced change in geometry blue-shifts the absorption by 0.10 eV ($\lambda_{\text{max}}^{(1)} \rightarrow \lambda_{\text{max}}^{(2)}$), which results in that the total net effect of the protein environment on the Q-band absorption is approximately 0.2 eV ($\lambda_{\text{max}}^{(1)} \rightarrow \lambda_{\text{max}}^{(3)}$).

Returning to the results in Table 5.2, it can be seen that all DFT-computed
Table 5.2. Calculated absorption maxima ($\lambda_{\text{max}}$, in eV) of the Q band of the BV chromophore in gas phase and protein environments

Reprinted from Paper I with permission from John Wiley & Sons. Copyright 2013 Wiley Periodicals, Inc.

<table>
<thead>
<tr>
<th>Functional</th>
<th>$\lambda_{\text{max}}^{(1)}$</th>
<th>$\lambda_{\text{max}}^{(2)}$</th>
<th>$\lambda_{\text{max}}^{(3)}$</th>
<th>$\lambda_{\text{max}}^{(1)}$</th>
<th>$\lambda_{\text{max}}^{(2)}$</th>
<th>$\lambda_{\text{max}}^{(3)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>BP86</td>
<td>1.84</td>
<td>1.94</td>
<td>1.91</td>
<td>0.10</td>
<td>0.07</td>
<td>-0.03</td>
</tr>
<tr>
<td>BLYP</td>
<td>1.83</td>
<td>1.93</td>
<td>1.90</td>
<td>0.10</td>
<td>0.07</td>
<td>-0.03</td>
</tr>
<tr>
<td>B97D</td>
<td>1.84</td>
<td>1.95</td>
<td>1.91</td>
<td>0.11</td>
<td>0.07</td>
<td>-0.04</td>
</tr>
<tr>
<td>$\tau$-HCTH</td>
<td>1.85</td>
<td>1.95</td>
<td>1.91</td>
<td>0.10</td>
<td>0.06</td>
<td>-0.04</td>
</tr>
<tr>
<td>B3LYP</td>
<td>1.86</td>
<td>1.96</td>
<td>2.05</td>
<td>0.10</td>
<td>0.19</td>
<td>0.09</td>
</tr>
<tr>
<td>PBE0</td>
<td>1.91</td>
<td>2.00</td>
<td>2.10</td>
<td>0.09</td>
<td>0.19</td>
<td>0.10</td>
</tr>
<tr>
<td>M06-HF</td>
<td>2.00</td>
<td>2.17</td>
<td>2.26</td>
<td>0.17</td>
<td>0.26</td>
<td>0.09</td>
</tr>
<tr>
<td>LC-BLYP</td>
<td>2.03</td>
<td>2.20</td>
<td>2.31</td>
<td>0.17</td>
<td>0.28</td>
<td>0.11</td>
</tr>
<tr>
<td>CAM-B3LYP</td>
<td>1.99</td>
<td>2.12</td>
<td>2.20</td>
<td>0.13</td>
<td>0.21</td>
<td>0.08</td>
</tr>
<tr>
<td>LC-\omega PBE</td>
<td>2.03</td>
<td>2.18</td>
<td>2.29</td>
<td>0.15</td>
<td>0.26</td>
<td>0.11</td>
</tr>
<tr>
<td>CIS(D)</td>
<td>1.94</td>
<td>2.05</td>
<td>2.19</td>
<td>0.11</td>
<td>0.25</td>
<td>0.14</td>
</tr>
<tr>
<td>Reference value</td>
<td>1.55f</td>
<td>1.65e</td>
<td>1.77f</td>
<td>0.10</td>
<td>0.22</td>
<td>0.12</td>
</tr>
</tbody>
</table>

Unless otherwise noted, all calculations carried out with the 6-31+G(d,p) basis set.

Environment = Gas: QM calculations carried out in the gas phase. Environment = Protein: QM/MM calculations carried out with the default QM system (QM1).

Geometry = Gas: Calculations based on the B3LYP/6-31G(d,p) optimized BV structure. Geometry = Protein: Calculations based on the QM/MM optimized (BV_{Full+Site}) structure.

CIS(D) absorption maxima calculated with the 6-31G(d,p) basis set and obtained as $S_0 \rightarrow S_1$ excitation energies.

SAC-CI-extrapolated value.

Experimental data from Ref. [33].

$\lambda_{\text{max}}^{(1)} \rightarrow \lambda_{\text{max}}^{(2)}$ values (0.09 – 0.17 eV) are in good agreement with the corresponding reference value (0.12 eV). Finally, it can also be seen that for the total protein-induced shift ($\lambda_{\text{max}}^{(1)} \rightarrow \lambda_{\text{max}}^{(3)}$), the reference value (~0.2 eV) is more accurately reproduced by the hybrid functionals (0.19 – 0.28 eV) than by the pure GGAs (0.06 – 0.07 eV).

To summarize the results in this paper, we can conclude that while the pure GGAs offer a better agreement with the experimental Q and Soret-band absorption maxima of DrBphP, the hybrid functionals better account for spectral shifts induced by the protein environment.
5.2 Paper II

To further investigate how the choice of density functionals affects computed UV-vis spectra of phytochromes, absorption and emission maxima in aqueous solution of four sterically locked chromophores (denoted 5Za15Ea, 5Ea15Ea, 5Zs15Za, 15Za) were calculated using a variety of density functionals.

The structures of the chromophores studied are shown in Fig. 5.3, with the C8 and C12 propionic chains had been replaced by hydrogen atoms. In this work, absorption maxima were estimated as vertical $S_0 \rightarrow S_1$ excitation energies and emission maxima as vertical $S_1 \rightarrow S_0$ emission energies. This approximation was, in a number of cases, validated by comparing $S_0 \rightarrow S_1$ excitation energies with the corresponding full UV-vis spectra obtained from tens of transition energies. Based on optimized geometries using B3LYP-PCM/SVP, absorption and emission maxima were obtained at the TD-DFT/cc-pVDZ level of theory in combination with a state-specific PCM description of the water solvent. Here, non-equilibrium solvation were used for absorption calculations, whereas equilibrium solvation were used for obtaining emission data.

We start by focusing on the results of the calculated absorption maxima. Overall, it is observed that all density functionals blue-shift the corresponding experimental Q-band absorption maxima. In this regard, it is also notable that a similar result is observed for the *D*BophP Q-band absorption maximum studied in Paper I.

Comparing the performance of the variety of functionals, it is found that the pure GGAs (BP86, BLYP, PBE and $\tau$-HCTH) more accurately reproduce the experimental value than the hybrid GGAs (BHHLYP, B3LYP and PBE0), followed by hybrid meta-GGAs (M06-2X and M06-HF) and long-range-corrected hybrid GGAs (LC-BLYP, CAM-B3LYP and $\omega$B97X-D). Specifically, while the pure GGA values deviate from the experimental values by $\sim 0.15$ eV, the values computed with the other functionals differ from the reference values by more than $\sim 0.3$ eV. For the emission maxima, in turn, it is found that also for these do the pure GGAs more accurately reproduce the experimental value than the other functionals.

While TD-DFT often yields reliable excitation energies, there are some well-documented deficiencies with this approach such as treatment of states with double-excitations character. To investigate if such effects are important for this set of chromophores, complementary calculations were carried out with the CASPT2 method. By analyzing the CASSCF wave function underlying the CASPT2 calculations, it is observed that such effects are not of importance for the $S_1$ states.

To further investigate the performance of the density functionals, it is also of interest to compare the calculated Stokes shifts (difference between absorption and emission maxima) with the corresponding experimental values. From this analysis, it is found that the pure GGAs (Stokes shifts in the 0.07 – 0.10 eV range) exhibit better agreement with the experimental Stokes shifts (0.04 – 0.07 eV) than B3LYP and PBE0 (0.12 – 0.16 eV). B3LYP and PBE0, in turn, offer better agreement with the experimental Stokes shifts than remaining hybrid functionals (0.18 – 0.25 eV). This finding together with the accuracy of the calculated absorption and emission maxima indicate that the well-known trend that hybrid functionals give more accurate excitation energies than pure functionals does not apply for bilin...
chromophores (at least as far as equilibrium geometries are concerned).
Figure 5.3. Structures of the locked bilin chromophores considered in this work (R = CH₂CH₂COOH). Reprinted from Paper II with permission from Elsevier. Copyright 2013 Elsevier B.V.
The characteristic glow of a firefly is created by a bioluminescent reaction inside the active pocket of the enzyme firefly luciferase. In this reaction, the enzyme catalyzes a conversion of the ground-state heterocyclic acid D-luciferin into the
chemiexcited \( S_1 \), first excited singlet state) oxyluciferin (OxyLH\(_2\)) light emitter, which subsequently emits light.

Depending on the conditions, OxyLH\(_2\) can exist in any of the six spectrally overlapping chemical forms shown in Fig. 6.1. There are two neutral forms: keto-OxyLH\(_2\) and enol-OxyLH\(_2\). The mono-anion, in turn, can exist as the phenolate-keto-OxyLH\(^-\), phenolate-enol-OxyLH\(^-\) or enolate-OxyLH\(^-\) form. Finally, double deprotonation gives the OxyL\(^{2-}\) di-anion.

To date, there is no consensus regarding the identity of the OxyLH\(_2\) light emitter. While both theoretical [35–37] and spectroscopic [38,39] studies have proposed that the phenolate-keto-OxyLH\(^-\) form is the light emitter, Naumov and co-workers [40] have recently reported experimental data favoring an enolate species (enolate-OxyLH\(^-\) or OxyL\(^{2-}\)).

In Paper III, the most probable form as the light emitter was predicted based on the intrinsic tendency of the light emitter to prefer a particular form of OxyLH\(_2\) in aqueous solution at pH 7. This was performed by calculating excited-state \( pK_E \) and \( pK_a \) values of all reactions shown in Fig. 6.1.

### 6.1 Paper III

![Figure 6.2. Structure of the phenolate-keto-OxyLH\(^-\)-water complex considered.](image)

Using a thermodynamic Born-Haber (BH) cycle, \( pK \) values in ground \( (pK^{BH}(S_0)) \) and excited states \( (pK^{BH}(S_1)) \) were obtained by calculating Gibbs free energies, in aqueous solution, of reactants (ketones/acids) and products (enols/bases) in the two states. Since it is easier to calculate accurate results for the difference between the \( pK(S_1) \) and \( pK(S_0) \) values than the corresponding \( pK(S_1) \) values, we believe that the most reliable procedure for obtaining the \( pK(S_1) \) values (denoted \( pK^{BH,corr}(S_1) \)) is to use the expression

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

(6.1)

Here, experimental \( pK^{exp}(S_0) \) values were taken from Ref. [41] or derived from calculated and experimental data. All calculations were carried out at the \( \omega B97XD/6-31+G(d,p) \) level of theory using a hybrid cluster-continuum model including...
11 explicit water molecules. The explicit water molecules were placed as shown in Fig. 6.2 for phenolate-keto-OxyLH\(^{-}\). The computed \(pK^{\text{BH,corr}}(S_1)\) and experimental \(pK^{\text{exp}}(S_0)\) are presented in Fig. 6.3.

![Figure 6.3. \(pK^{\text{exp}}(S_0)\) and \(pK^{\text{BH,corr}}(S_1)\) values of OxyLH\(_2\).](image)

Starting by comparing keto-OxyLH\(_2\) with enol-OxyLH\(_2\) (reaction I in Fig. 6.3), it is seen that while the neutral forms are of similar stability (\(pK^{\text{exp}}(S_0) = -0.39\)) in the ground state, the calculated \(pK^{\text{BH,corr}}(S_1)\) value of \(\sim 5\) suggests that keto-OxyLH\(_2\) is the dominating neutral form in the excited state. In Fig. 6.3, it can also been seen that keto-OxyLH\(_2\) is a strong photoacid with a calculated \(pK^{\text{BH,corr}}(S_1)\) value of about 2 (reaction IV). This value indicates that phenolate-keto-OxyLH\(^{-}\) is more stable than keto-OxyLH\(_2\) in the excited state. Therefore, we can conclude that neither keto-OxyLH\(_2\) nor enol-OxyLH\(_2\) is expected to significantly contribute to the light emission.

Having shown that the neutral forms are likely of minor importance for the light emission, we will now investigate to what extent the three mono-anionic forms contribute to the emission. As for the phenolate-keto-OxyLH\(^{-}\) = enolate-OxyLH\(^{-}\) equilibrium (reaction II in Fig. 6.3, which corresponds to a keto-enol tautomerization and subsequent internal proton transfer from the enolic hydroxyl group to the corresponding phenolic group), it can be observed that despite that enolate-OxyLH\(^{-}\) predominates in the ground state, our calculated \(pK^{\text{BH,corr}}(S_1)\)
value (~6) suggest that it is actually phenolate-keto-OxyLH\(^-\) that is more stable in the excited state. For the phenolate-keto-OxyLH\(^-\) = phenolate-enol-OxyLH\(^-\) equilibrium (reaction III in Fig. 6.3), in turn, we also note that the calculated pK\(_{E}^{BH,corr}(S_{1})\) value (~7) indicates that phenolate-keto-OxyLH\(^-\) is more stable in the S\(_{1}\) state.

Based on the analysis above, we can conclude that either phenolate-keto-OxyLH\(^-\) or OxyL\(^2\)(\(^-\)) is the most probable emitter. As shown in Fig. 6.3 (reaction IV), the calculated pK\(_{E}^{BH,corr}(S_{1})\) value (~14) suggests that phenolate-keto-OxyLH\(^-\) is more stable than OxyL\(^2\)(\(^-\)) upon excitation to the S\(_{1}\) state.

To summarize the results, we can therefore conclude that phenolate-keto-OxyLH\(^-\) is intrinsically the most stable form of OxyLH\(_2\) in the excited state. This finding gives further support for the hypothesis that phenolate-keto-OxyLH\(^-\) is the light emitter of firefly.
Conclusions

This thesis is devoted to theoretical studies aimed at contributing to the understanding of photochemical reactions in phytochromes and firefly luciferase. It has been shown that pure GGAs more accurately reproduce experimental absorption and emission maxima of bilin chromophores than hybrid functionals, whereas hybrid functionals better account for spectral shifts induced by apophytochromes. For firefly luciferase, in turn, it has also been shown that the phenolate-keto-OxyLH\textsuperscript{−} form is intrinsically the most stable form of OxyLH\textsubscript{2} in the excited state. Although some new findings have been gained for these systems, there are still many questions to be resolved.

As for phytochromes, a future work would aim to give a more detailed description of the \textit{in vivo} excited-state isomerization mechanism of the bilin chromophore, which is the key photochemical problem today in this field. Although a number of theoretical studies have been concerned with the excited-state isomerization in the gas phase or in solution, the impact of specific interactions between the chromophore and the surrounding apoprotein on this reaction have not been accounted for. In this regard, the results presented in this thesis can contribute to a more critical assessment of the methods available today. Another interesting topic for further studies is to investigate how the absorption and emission maxima can be tuned by either modifying the chromophore or the apoprotein.

For firefly luciferase, in turn, the next step is to investigate the excited-state equilibria of OxyLH\textsubscript{2} in the natural environment inside the active pocket. Although the intrinsic stabilities of the various forms of OxyLH\textsubscript{2} studied in this work give some useful information on the \textit{in vivo} excited-state equilibria, the surrounding environment is likely to impact on the individual equilibrium reactions.

Finally, it is my hope that the results presented in this thesis will not only contribute to a deeper understanding of the photochemistry of phytochrome and firefly luciferase chromophores but also have a bearing beyond these specific sys-
tems in other photochemical studies.


Publications included in this thesis:

I Modeling of phytochrome absorption spectra
O. Falklöf, B. Durbeej

II Red-light absorption and fluorescence of phytochrome chromophores: A comparative theoretical study
O. Falklöf, B. Durbeej

III Deciphering excited-state protonation and tautomeric states of oxyluciferin through calculation of excited-state equilibrium constants
O. Falklöf, B. Durbeej
Manuscript

Related work:

IV On the origin and variation of colours in lobster carapace
S. Begum, M. Cianci, B. Durbeej, O. Falklöf, J. R. Helliwell, A. Hädener, A. Regan, I. Watt
Manuscript
Publications

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-103338