

Quantum Chemical Investigations on Sensory Photoreceptors in Natural and Artificial Systems



Dissertation

zur Erlangung des Doktorgrades der Naturwissenschaften (Dr. rer. nat.)
der Fakultät IV
- Chemie und Pharmazie -
der Universität Regensburg

vorgelegt von

Keyarash Sadeghian

aus Teheran/Iran

Regensburg 2008

Promotionsgesuch eingereicht am:	05.11.2008
Tag des kolloquiums:	19.12.2008
Diese Arbeit wurde eingeleitet von:	Prof. Dr. Martin Schütz

Promotionsausschuss

Vorsitzender:	Prof. Dr. Georg Schmeer
Erstgutachter:	Prof. Dr. Martin Schütz
Zweitgutachter:	Prof. Dr. Bernhard Dick
Drittprüfer:	Prof. Dr. Jörg Daub

Die Ergebnisse dieser Arbeit sind bereits veröffentlicht worden:

Chapter 3

K. Sadeghian and M. Schütz

"On the photophysics of artificial blue-light photoreceptors:
an ab initio study on a flavin-based dye dyad at the level of
coupled-cluster response theory"

Journal of the American Chemical Society

129, 4068-4074 (2007), doi: 10.1021/ja068536t

Chapter 4

K. Sadeghian, M. Bocola and M. Schütz

"A Conclusive Mechanism of the Photoinduced Reaction Cascade in Blue
Light Using Flavin Photoreceptors"

Journal of the American Chemical Society

130, 12501-12513 (2008), doi: 10.1021/ja803726a

Chapter 5

J. Svoboda, B. König, K. Sadeghian and M. Schütz

"2'-Oxoethyl Flavin Revisited"

Zeitschrift für Naturforschung: B

63, 47-54 (2008)

Acknowledgements

First of all I would like to thank my supervisor, Prof. Dr. Martin Schütz, for his full support over the past four years. I also thank him for his trust in giving me the opportunity to enter the field of quantum chemistry. I also wish to thank Dr. Marco Bocola who has been a tremendous help in setting up the QM/MM calculations and for giving me some insight into the field of computational biology.

I have enjoyed working in the theoretical chemistry group of the University of Regensburg and therefore thank my kind and supportive colleagues, Dr. Denis Usvyat, Danylo Kats, Dominik Schemmel and Marco Lorenz. The efforts of our system administrator, Klaus Ziereis, is specially appreciated at this stage.

Within the Graduate College "Sensory Photoreceptors in Natural and Artificial Systems", I have mostly benefited from discussions with Prof. Jörg Daub, Prof. Alfons Penzkofer and Prof. Bernhard Dick. I have also enjoyed the discussions with my friends and countrymen, Dr. Peyman Zirak and Dr. Javid Shirdel.

I wish to express my deep gratitude to my parents, Siavash Sadeghian, Farzaneh Fathnezhad and my sister Ladan Sadeghian, who have been supporting and motivating me throughout my studies over the years. My uncle, Djahangir Sadeghian, has always offered his help and support when I needed it the most and I am therefore very grateful for having him.

Last but not least, I want to thank my wife, Sharareh Arabbagheri, for her continuous support, motivation, understanding and encouragement during the difficult times.

Contents

1	Blue Light Photoreceptors	3
1.1	Biological Photoreceptors	3
1.2	Cryptochromes (CRY)	5
1.3	CRY-type Artificial Photoreceptor	7
1.4	BLUF Domains	9
2	QM/MM methodes	17
2.1	Concept	17
2.2	Preparation of QM/MM input structures	25
2.3	ChemShell interface	31
2.4	QM/MM Geometry optimization setup	31
2.5	Choice of QM region	33
2.6	Choice of QM theory: Ground state	34
2.7	Choice of QM theory: Excited state	36
3	Phenothiazine-Flavin Dyad	39
3.1	Introduction	39
3.2	Computational details	41
3.3	Results and Discussion	43
3.3.1	At the ground state geometry	43
3.3.2	At the CT minimum-energy geometry	45
3.3.3	At the conical intersection structure	48
3.3.4	At the "LE-minimum energy" geometry	49
3.3.5	Solvent Effects	50
3.4	Conclusions	53
4	BLUF DOMAINS	54
4.1	Introduction	54
4.2	Computational Details	57
4.3	Results and Discussion	60
4.3.1	Search for a representative structure	60

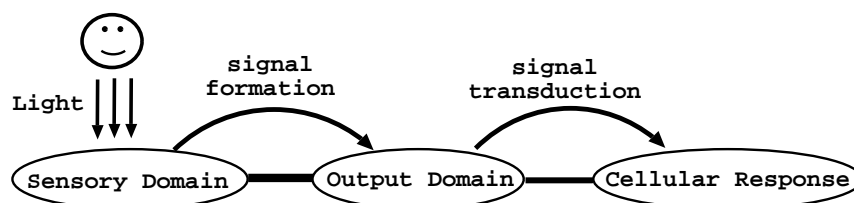
<i>CONTENTS</i>	2
4.3.2 Ground state QM/MM: Determination of the most stable dark state structure	63
4.3.3 Choice of QM method for the excited states	65
4.3.4 Charge transfer QM/MM: Biradical formation mechanism	66
4.3.5 Open-shell QM/MM: Biradical recombination path	69
4.3.6 Path to the Signaling State	71
4.3.7 Vertical excitation energies: comparison of dark, Q51- <i>E</i> and Q51- <i>Z</i> structures	74
4.3.8 QM/MM vibrational frequency calculation	78
4.4 Conclusions	79
5 2'-Oxoethyl Flavin Revisited	81
6 Summary	86
Bibliography	88

Chapter 1

Blue Light Photoreceptors

1.1 Biological Photoreceptors

The Earth's surface has been exposed to the solar radiation during the course of evolution. Living organisms have therefore found ways to adapt themselves by developing molecular machineries, enabling them to detect, respond to and in some cases even use the absorbed photon of light as a source of energy for their metabolism. A perfect example is the cyanobacteria family of prokaryotes with their sensory proteins which allow them to respond to the UV and/or visible range of the solar spectrum. The photosensory/photoregulatory parts of such domains are responsible for the light detection, whereas the signal transduction, a precursor step for the final cellular response, involves output domains (see Fig. 1.1)[1, 2]. The photosensory domains/proteins, being one center



of attention in the Graduate College GRK-640 (Sensory photoreceptors in natural and artificial systems), can be divided into six important families[3] (see Fig. 1.1). The rhodopsin[4], Phytochrome[5] and Xanthopsins[6] families have different chromophores, and undergo isomerization of the chromophore as part of their light perception cascade. The second category of the regulatory domains are the blue light photoreceptors: Light-Oxygen-Voltage (LOV) domains[7, 8], Cryptochromes (CRY)[9] and Blue-Light-Using-Flavin (BLUF)[10] domains, all having flavin as their active chro-

mophore, but differing in the signaling state formation.

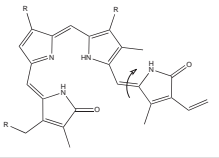
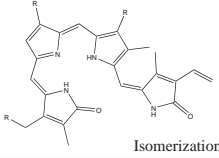
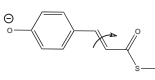
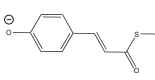
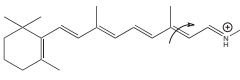
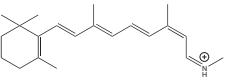
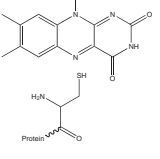
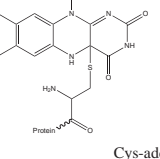
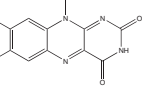
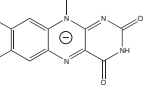
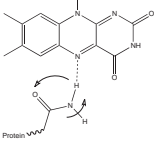
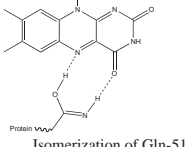
Domain	Dark state	Signaling state	Chromophore
Phytochrome		 Isomerization	linear tetrapyrrole
Xanthopsin		 Isomerization	<i>p</i> -coumaric acid
Rhodospin		 Isomerization	retinal
LOV		 Cys-adduct	flavin
CRY		 Charge Transfer	
BLUF		 Isomerization of Gln-51	

Figure 1.1: Different families of photoreceptors

In this project, the photophysics of an artificial photoreceptor of cryptochrome type and the signaling state formation in BLUF domain proteins were studied. These two blue light photoreceptors will be briefly introduced in the following sections.

1.2 Cryptochromes (CRY)

Being the oldest member of the blue light photoreceptor family, Cryptochrome has been found in prokaryote (*Cyanobacteria*), low and higher eukaryote such as plants (*Arabidopsis*) and insects (*Drosophila*)[11]. Synchronization of the circadian clock in animals, seed germination and pigment accumulation in plants are some of the functions (cellular response) which has been associated with these sensory domains.

The crystal structure of Cry-3[12], a cryptochrome found in *Arabidopsis Thaliana*, shown in Fig.1.2(a), reveals the light-harvesting antenna, Methenyltetrahydrofolate (MTHF), and the catalytic cofactor, Flavin Adenine Dinucleotide (FAD), both being non-covalently bound to protein. A similar situation is encountered in the evolutionary related photolyase proteins, which are responsible for photo-reactivation/repair of UV-induced damages of pyrimidine bases in DNA. The so called cyclobutane pyrimidine dimers (CPD) are formed in a $[2\pi + 2\pi]$ cycloaddition of adjacent

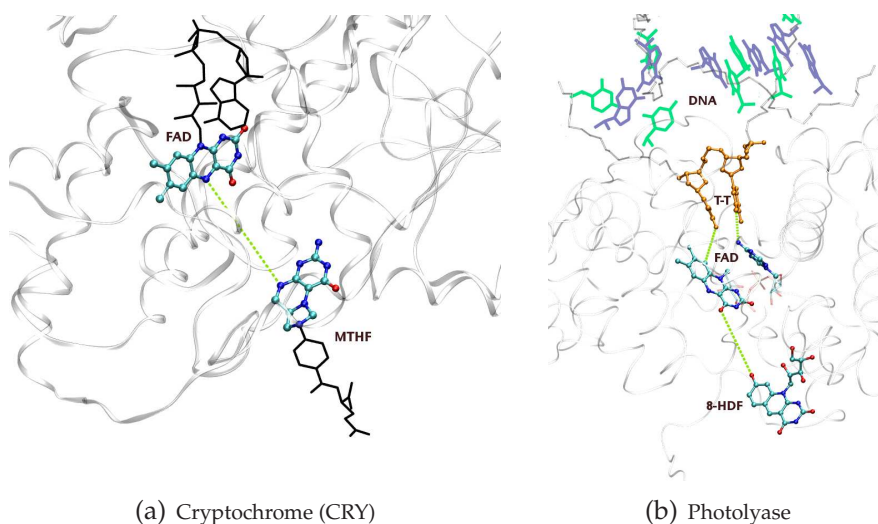


Figure 1.2: Crystal structure of Cry-3[12], where the flavin (redox mediator) and the MTHF (antenna) are highlighted. Crystal structure of photolyase bound to DNA after CPD repair[13]. The two thymine bases are shown in orange.

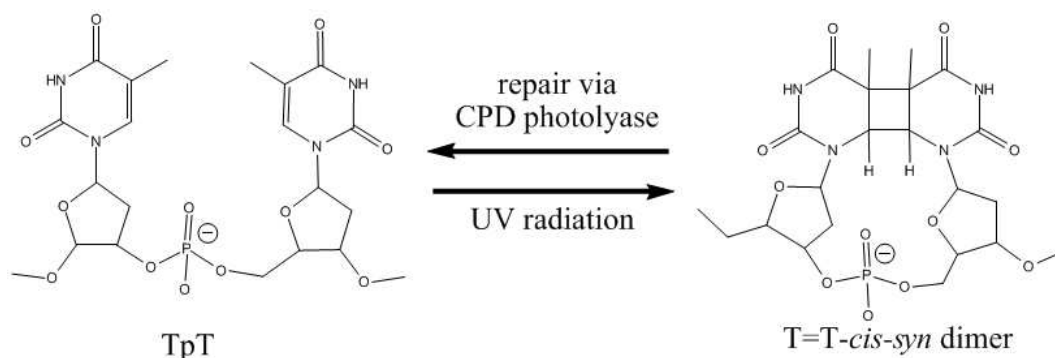


Figure 1.3: UV-induced Formation and repair of the CPD lesion.[15].

pyrimidine bases, mainly Thymines ($\langle \text{TT} \rangle$) [14] (see Fig. 1.3), upon UV exposure.

The photolyase and cryptochrome proteins are structurally similar but no common biological function is known to this date. Only recently it was found that cryptochrome may have photolyase activity for single stranded DNA containing CPD, hence stressing the strong link between these two protein families [16]. Although the mechanism with which cryptochromes transform the absorbed photon of (blue) light into the desired biological signal is not fully known, but based on structural similarity with the photolyase, it is assumed that the signaling state formation cascade in the two proteins involves similar initial photochemical steps.

In photolyase the repair mechanism is triggered by light which is absorbed by an antenna pigment such as 8-hydroxy-5-deazaflavin (8-HDF) (deazaflavin class of photolyase) or Methenyltetrahydrofolate (MTHF) (folate class) [17]. The absorbed energy is then transferred to the second chromophore, namely the reduced flavin coenzyme (FADH^-) which finally acts as an electron donor to the CPD thereby initiating the repair process [18]. The MTHF or 8-HDF chromophores are not necessarily needed for the repair process, but due to their much higher extinction coefficient and absorption at higher wavelengths can increase the repair rate by 1-2 orders of magnitude at dim light conditions [19]. The FADH^- chromophore on the other hand, is absolutely essential for both DNA-binding, a process which can also occur in the dark [20], and the light-dependent repair process itself [21].

1.3 CRY-type Artificial Photoreceptor

The cryptochrome and photolyase proteins have proved to be very efficient in their function, hence attempts have been made to learn from the interactions between their functional units (i.e. their chromophores). The goal is to synthesize an artificial photoreceptor which can eventually be *tailored* for a given application. Prof. Daub and coworkers have been quite active in this field and as part of the Graduate College GRK-640 synthesized and studied a molecular arrangement, schematically shown in Fig.1.4(a), aimed at mimicking the function of the cryptochrome/photolyase blue light photoreceptor[22].

The chromophores were chosen in such a way, that their intramolecular interactions, i.e. charge transfer and energy transfer, resemble those found in the cryptochrome/photolyase. First of all an efficient light-harvesting antenna is required, for which the pyrene molecule, due to its high extinction coefficient at 340nm and emission peak at 380-400nm, was chosen. This subunit corresponds to the MTHF-moiety in cryptochrome and its emission indeed covers the absorption band of flavin[23]. The isoalloxazine core of FAD in cryptochrome/photolyase was used as the redox mediator in the whole system and finally a phenothiazine moiety was chosen to be the electron donor.

These three dyes are covalently bound together, using linking moieties, forming a pyrene-flavin-phenothiazine triad as shown in Fig.1.4(b). This arrangement together with the phenothiazine-flavin, phenothiazine-pyrene and pyrene-flavin dyads were synthesized¹ and further investigated via fluorescence measurements[24, 25, 26, 27], spectroelectrochemistry experiments[28, 29, 30] and quantum chemical (semi-empirical) calculations[31, 32].

We decided to investigate the electron donor-acceptor subunit, i.e. the phenothiazine-flavin dyad, of this molecular arrangement depicted in Fig. 1.5. Spectroelectrochemical measurements indicate the formation of radical cation and anions[30] which is a sign for photo-induced electron transfer in such donor-acceptor systems. A further indication for such a process is provided by time-resolved fluorescence measurements[30, 24] where low quantum yield and fluorescence quenching were detected.

Quantum chemical calculations, performed as part of this project, deal with the photophysics of such a dyad and were aimed at explaining the experimental results by employing *ab initio* methods, such as TD-DFT

¹Roman Procházka, Dissertation, 2004, University of Regensburg

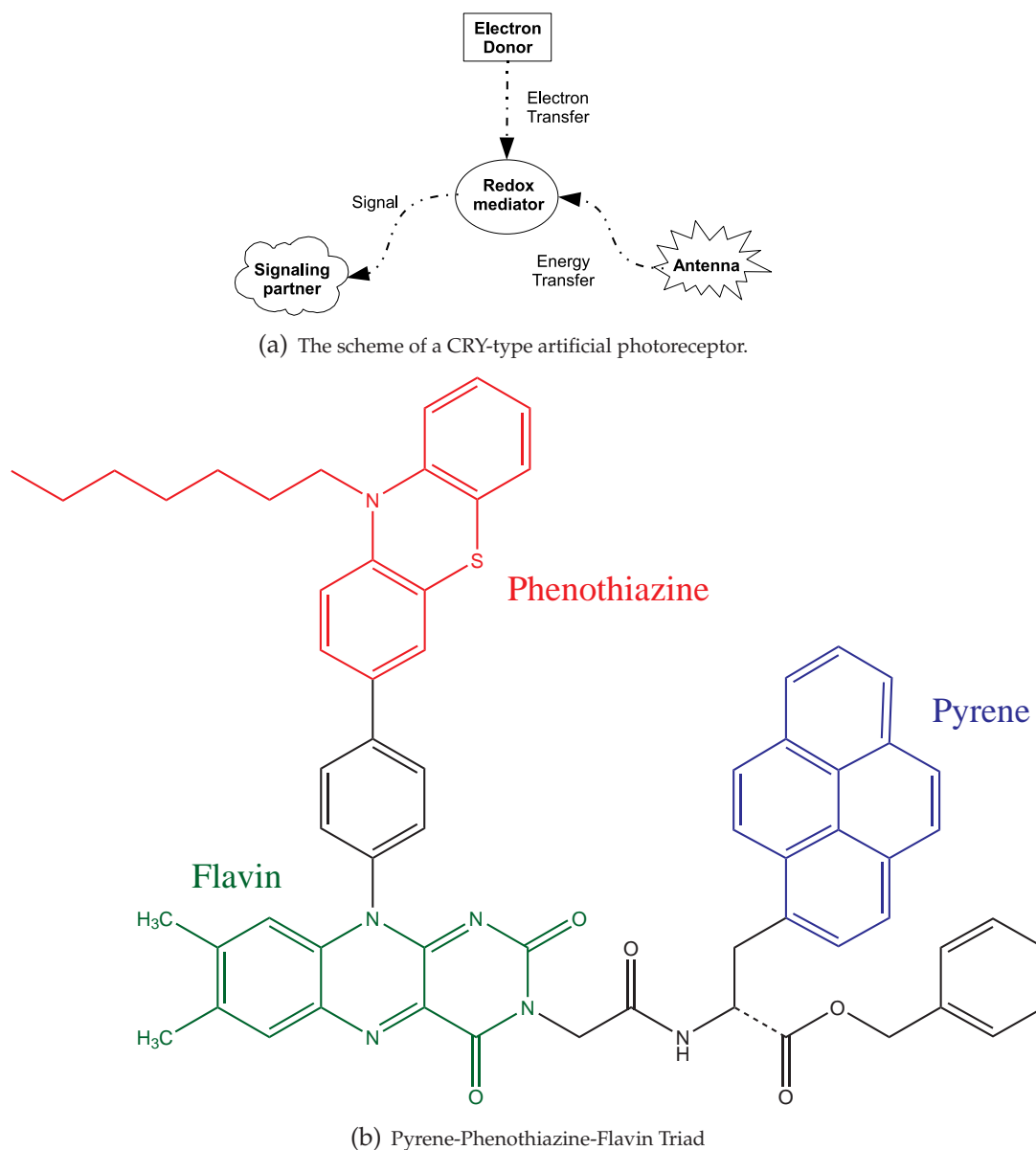


Figure 1.4: A molecular system aimed at mimicking the function of cryptochrome (adapted from Ref. [22])

and TD-CC2. These methods are used to pick out those low lying excited states which of special interest. Apart from geometry relaxation (vibrational cooling) which is expected to occur spontaneously upon light absorption, competing radiative (fluorescence, phosphoresence) and non-radiative (intersystem crossing, conical intersections between ground and excited states) decay processes also play a role[33].

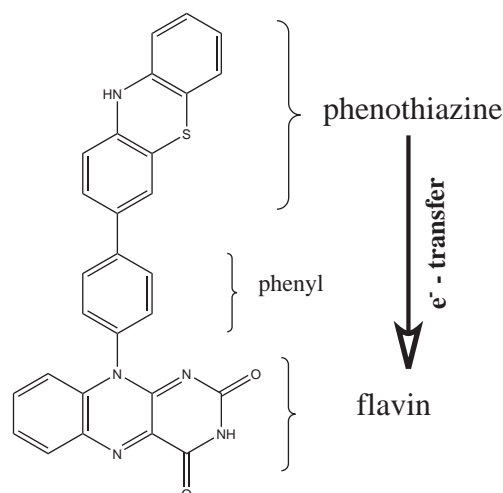


Figure 1.5: phenothiazine-pyrene-flavin molecule

In this special system, the locally excited (LE) state of flavin and the phenothiazine $\xrightarrow{e^-}$ flavin charge transfer (CT) state, have been shown to play an important role. Using the above mentioned quantum chemical methods, we propose a mechanism for the underlying photoinduced cascade of events in this dyad which also explains the experimental results (see chapter 3). Here the charge transfer, due to its low oscillator strength, is not directly populated via excitation. A conical intersection between the initially populated LE state and the CT state, is responsible for the deactivation of the former and leads to the experimentally observed low fluorescence quantum yield.

A conical intersection mediated photoinduced electron transfer does not seem to be a rare thing in nature and we have observed a similar situation in the signaling state formation of BLUF domains which will be described further below.

1.4 BLUF Domains

In the second part of the project, we studied a biological photoreceptor family, namely the Blue Light Using FAD (BLUF) domain[10]. This domain was first discovered in AppA[36], a multi-domain protein from the phototropic bacterium *Rhodobacter sphaeroides*. The N-terminus domain of AppA was later found to regulate photosynthetic gene expression by sensing blue light and/or redox conditions and interacting with a second (repressor) protein PpsR[37, 38, 39]. Another example for a BLUF domain

with a biologically known function is photo-activated adenylyl cyclase (PAC) from *Euglena gracilis* which is involved in the blue-light dependent control of enzyme activity[40].

The underlying mechanism of the photo-induced signaling state formation in BLUF domains has been a topic under discussion ever since the crystal and NMR structures were published in the years 2005-06 (see Tab.1.1).

Table 1.1: Crystal/NMR structures resolved and published to this date.

Name	Tll0078	AppA	BlrB	AppA 5-125	AppA C20-S	Slr1694
Exp. Method	X-ray	X-ray	X-ray	NMR	X-ray	X-ray
Resolution (Å)	2.00	2.30	1.90	-	2.30	1.80
Release Date	7.6.'05	28.6.'05	24.7.'05	7.12.'05	6.9.'06	19.12.'06
PDB code	1X0P	1YRX	2BYC	2BUN	2IYG	2HFN

In all known natural photoreceptors (Fig. 1.1), the signal generation is accompanied by a change in the structure of the corresponding sensory domain and BLUF proteins are no exception. In contrary to, say Rhodopsins,

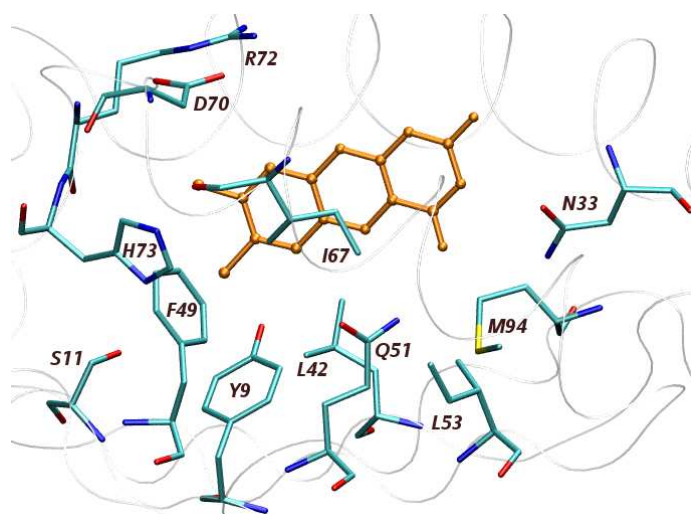


Figure 1.6: Crystal structure of BlrB BLUF domain[34]. The flavin chromophore (orange) together with the highly conserved residues are also shown. Of great importance for the photocycle are the Tyr-9, Gln-51, Asn-33 and Met-94 residues. The Trp-94 residue (not strictly conserved, therefore not shown) has been proposed to swap its position with the Met-92 moiety[35]. See text for further detail.

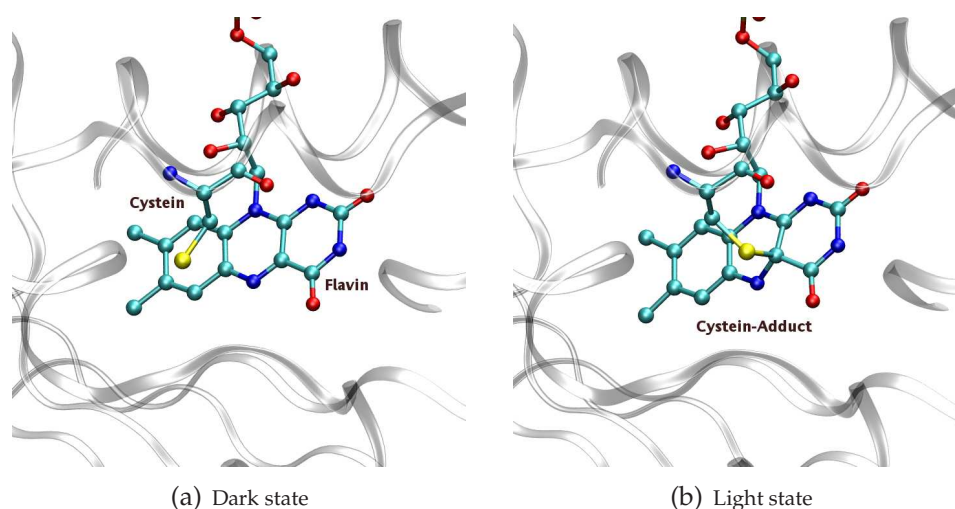


Figure 1.7: Crystal structure of LOV2 domain from the phototropin photoreceptor in the dark[41] and light/signaling [41] states

the non-covalently bound flavin chromophore is clearly unable to undergo any *cis-trans* isomerization. The anticipated structure change must therefore involve flavin and more importantly its neighboring highly conserved residues (shown in Fig. 1.6). In the case of LOV domains, another class of blue light photoreceptors, a photo-adduct (the so called cys-adduct) with a distinct stable structure and spectroscopical signatures is formed (cf. Ref [42] and references therein). The signaling state formation cascade involves an intersystem crossing process between the initially populated long living singlet excited state and the first triplet excited state. The flavin neighboring and highly conserved cysteine residue (in LOV domains) reacts with the chromophore in the triplet state leading to the cys-adduct (see Fig. 1.7). This initiates the change in the protein conformation and subsequently leads to the biological signal formation and its associated cellular response like phototropism[43].

In the case of BLUF domains however the situation is completely different to that in LOV domains and no covalent bond between flavin and any of the neighboring residues is formed². The characteristic signatures for the signaling state structure and its formation in BLUF domains are the

²Recently, Suzuki *et. al.* mutated the highly conserved ILE66 residue of a BLUF domain (cf. Fig. 1.6) with a cysteine amino acid. From the structural point of view the ILE-66 residue corresponds more or less to the Cys residue in LOV domains as they are both positioned on top of the flavin in direct contact with the N5 and C4a atoms (compare with Fig. 1.7). Interestingly they were able to show that such a mutation in a BLUF domain lead to similar cys-adduct as found in LOV domains[44].

following (see Ref.[45] and references therein):

- (i) A red-shift of ca. 10nm in the S_0 - S_1 absorption band which corresponds to a local excitation within the flavin chromophore.
- (ii) A red-shift of ca. 10 wave numbers in the IR band around 1700 cm^{-1} which has been attributed to the stretch mode of the $C4=O4$ bond of flavin.
- (iii) Appearance of a transient IR absorption-band around 1666 cm^{-1} . This mode is only observed in the light-adapted form of the BLUF domain and thereafter suggested to be a marker for the signaling state conformation[46].
- (iv) The rate for the signaling \rightarrow dark state transition (the dark state recovery rate) is in the order of some minutes. Compared with the much faster rate for the signaling state formation (c.f. Fig. 1.8), we are faced with a subtle photo-induced change in the structure of flavin's neighborhood. These changes can be reversed, in the absence of light, via thermal processes, which obviously occur on a much longer time scale.

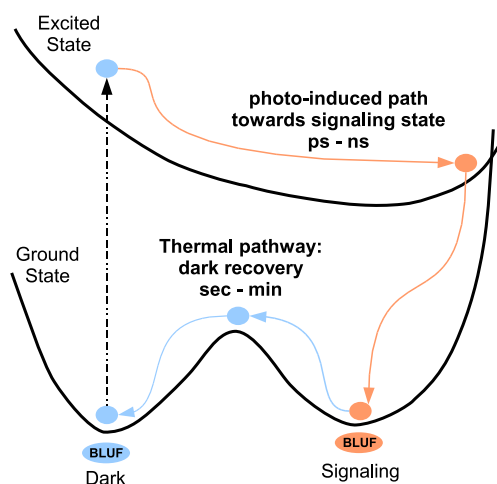


Figure 1.8: The photocycle of BLUF domains shown schematically

By employing a QM/MM scheme to study both ground and excited state conformations of BLUF domain's active site, we have tried to investigate the following issues and come up with a possible mechanism:

- The orientation of the glutamine (Gln-51) residue. This has been the matter of debate and should be clarified with the help of accurate calculations. Unfortunately, the crystal structures published to this date (Tab.1.1) do not deliver any clear picture, the reason being that the amide (NH_2) and carbonyl ($\text{C}=\text{O}$) groups of the Gln-51 can not be resolved in the presently available structures. Hence, in combination with different possible orientations for the neighboring Tyr-9 residue, we are faced with three choices for the hydrogen bonding network as shown in Fig. 1.9.

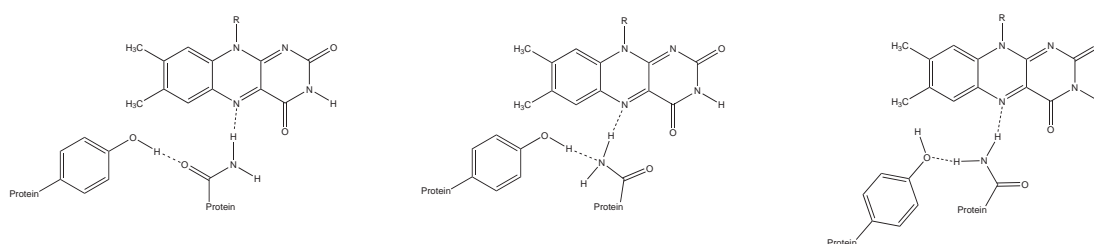
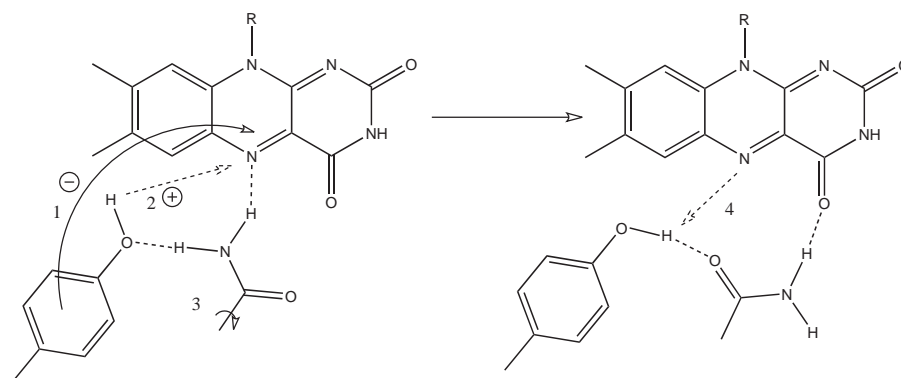


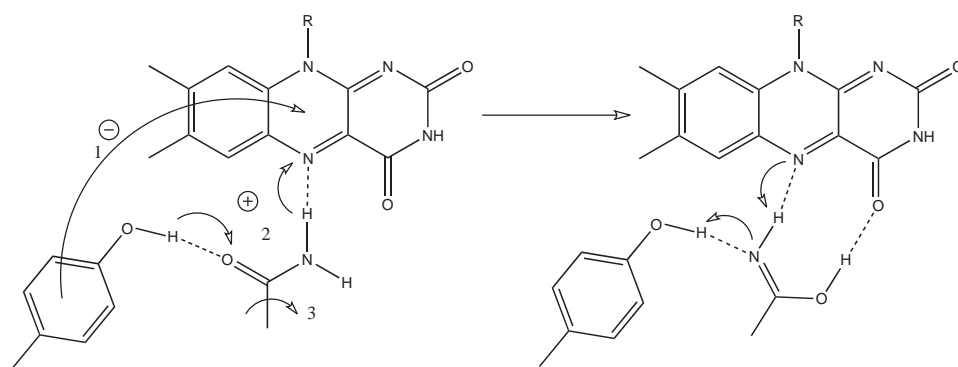
Figure 1.9: Possible orientations of the Gln-51 and Tyr-9 residues towards the flavin chromophore.

- Photoinduced Processes: A charge transfer (CT) state is proposed to be involved in the photochemistry of BLUF domains. Thereafter it is desirable to gain adequate knowledge about the energetic position of such excited states. Since the flavin moiety is expected to be locally excited (LE state) after light exposure, there must be a way of activating the charge transfer (CT) state³. In the case of phenothiazine-flavin dyad, a conical intersection between the two LE and CT states was found to be the key step, hence explaining the experimental result. In a direct comparison with the case of BLUF domains we can assign the role of an electron-acceptor to flavin and an electron-donor to the Tyr9 (phenothiazine in the dyad molecule) and therefore assume/propose the similar photophysics as shown for the dyad system. Furthermore by calculating the vertical excitation energies, we can consider/rule out other electron donors in BLUF-domains, e.g. the Trp-92. To this end, it is crucial that (i) Tyr-9, Gln-51 are treated quantum mechanically (\rightarrow QM-region), (ii) the remaining residues of the proteins are also included in the calculation (as point charges \rightarrow MM-region).

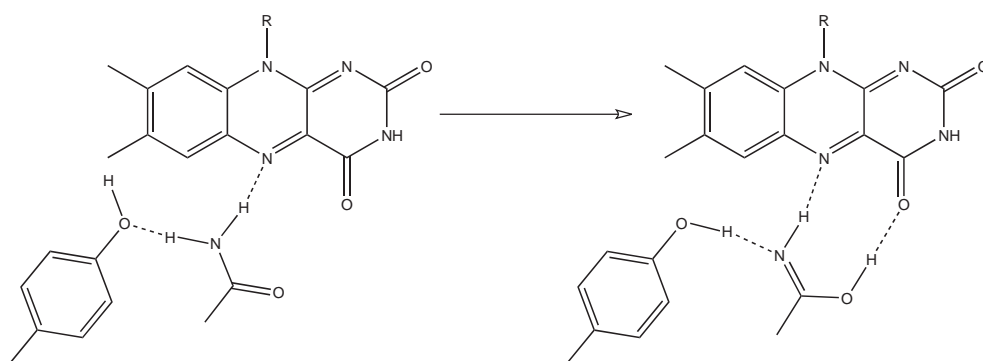
³Charge transfer states usually have very low oscillator strengths (\rightarrow absorption coef-



(a) Mechanism proposed by Gauden *et. al.* involves rotation of the Gln-51 moiety but no isomerization of the latter residue[47]



(b) Mechanism proposed by Domratheva *et. al.* involves the rotation and isomerization of the Gln-51 residue[48]



(c) Stelling *et. al.* proposed the isomerization of Gln-51 without any rotation[46]. The authors assume a Trp-in conformation (see Text) for the dark state.

Figure 1.10: Different mechanisms proposed for the signaling state formation in BLUF domains

ficient) and are therefore not directly accessible via photoexcitation

- A short-lived flavin radical (presumably in combination with a Tyr9-radical) has been identified experimentally. The formation of such biradical species will be looked at in more detail in the context of the QM/MM calculations presented in this work. One of the key questions which can also be answered is the radical recombination process. Experimentally, a rather short lifetime for this process is measured/estimated and one would expect a fast pathway for the latter process.

Several pathways for the formation and recombination of the mentioned $\text{Tyr}^\bullet\text{-FAD}^\bullet$ have been suggested in the literature (see Fig. 1.10). These include the rotation of the Gln-51 in its amine form by Gauden *et. al.* (see Fig. 1.10(a)) or in its imidic form by Domratcheva *et. al.* (see Fig. 1.10(b)). Meech *et. al.* also suggested the isomerization of the Gln-51 moiety (see Fig. 1.10(c)). In contrast to the first two groups these authors proposed that the Gln-51 moiety may not be rotated during the signaling state formation. For the dark state structure however, they suggest a Trp-in conformation (see below). In the results section, we show that a direct pathway for the radical recombination is in a good agreement with the experimental results available.

- One further issue regarding the dark and signaling states of BLUF domains is the role of the (not strictly conserved) Trp-92 moiety. In some of the published structures, one finds this residue to swap its position with the highly conserved Met-94 residue in the flavin binding pocket (see Fig. 1.11). A switch between the two Trp-in and Met-in conformations, which most probably influences the interaction of the BLUF domain with the output domain (e.g. AppA \rightarrow PpsR[39], has been suggested to be the main biological signal[35]. Possible stabilization of the signaling state due this switch is also studied and will be presented.

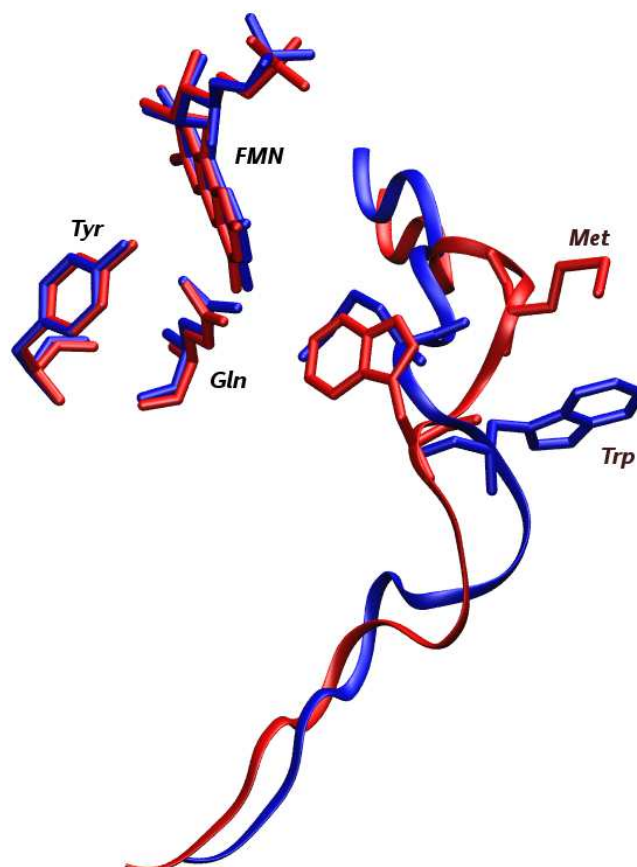


Figure 1.11: Superposition of the Met-in structure (blue) from BlrB[34] (in green) with the Trp-in structure (red) of AppA[49] (in magenta) BLUF domains. The Trp-Met switch is accompanied by the change in the conformation which can be clearly seen from the twist caused in the protein backbone in the vicinity of the these residues.

Chapter 2

QM/MM methodes

2.1 Concept

Biopolymers such as proteins or DNA are typical systems of interest when we study important biological processes. These entangled long chains of covalently bound nucleic or amino acids, in combination with their native water environment, create a rather complex system consisting of many thousands of atoms.

Quantum Mechanics (QM)

Of course it is desirable to treat all these atoms with the laws of quantum mechanics (QM). This means defining a wave function Ψ as a function of time (t), electron (r) and nuclear (R) coordinates, from which any property can be extracted after solving the time-dependent Schrödinger equation:

$$H \Psi(r, R, t) = i \frac{\partial}{\partial t} \Psi(r, R, t)$$

For a time-independent Hamilton operators, this equation is written in the form,

$$H \Psi(r, R) = E_T \Psi(r, R) \quad (2.1)$$

with E_T being the total energy of the system. H is the *total* Hamilton operator consisting of kinetic and potential energy terms for the electrons and nuclei of the system:

$$H = T_n + T_e + V_{ne} + V_{ee} + V_{nn} \quad (2.2)$$

The first two terms, T_n and T_e , represent the kinetic energy of the nuclei and electrons respectively. The V_{nn} , V_{ee} and V_{ne} terms correspond to the nucleus-nucleus, electron-electron (both repulsive) and electron-nuclei (attractive) interactions respectively. We may extract an *electronic* Hamilton operator, H_e , by subtracting the kinetic energy of the nuclei from H :

$$H_e = T_e + V_{ne} + V_{ee} + V_{nn} \quad \rightarrow \quad H = T_n + H_e \quad (2.3)$$

We now assume, that we have the full set of solutions, $\psi_i(r, R)$ ($i = 1, 2, 3, \dots, \infty$), to the electronic Schrödinger equation:

$$H_e(R) \psi_i(r, R) = \epsilon_i(R) \psi_i(r, R) \quad (2.4)$$

This allows us to write the *exact* wave function, $\Psi(r, R)$, as an expansion of the eigenfunctions, $\psi_i(r, R)$, of the electronic Hamilton operator H_e (see for example Ref.[50]):

$$\Psi(r, R) = \sum_{i=1}^{\infty} \chi_{ni}(R) \psi_i(r, R) \quad (2.5)$$

Here, the expansion coefficients, $\chi_{ni}(R)$, are functions of the nuclear coordinates only. By plugging this ansatz in the time-independent Schrödinger equation, Eq.(2.1), we obtain¹:

$$\begin{aligned} \sum_{i=1}^{\infty} (\nabla_n^2 + H_e) \chi_{ni} \psi_i &= E_T \sum_{i=1}^{\infty} \chi_{ni} \psi_i \\ \sum_{i=1}^{\infty} \nabla_n^2 \chi_{ni} \psi_i + \sum_{i=1}^{\infty} \chi_{ni} H_e \psi_i &= E_T \sum_{i=1}^{\infty} \chi_{ni} \psi_i \\ \sum_{i=1}^{\infty} \nabla_n^2 \chi_{ni} \psi_i + \sum_{i=1}^{\infty} \epsilon_i \chi_{ni} \psi_i &= E_T \sum_{i=1}^{\infty} \chi_{ni} \psi_i \end{aligned}$$

¹the explicit form of the kinetic energy operator,

$$T_n = \sum_a^{N_n} -\frac{1}{2M_a} \nabla_a^2$$

has been redefined, as ∇_n^2 for simplicity.

We can now use the orthonormality condition for the eigenfunction H_e operator² to obtain[50]

$$\sum_{i=1}^{\infty} \left\{ 2\langle \psi_j | \nabla_n | \psi_i \rangle (\nabla_n \chi_{ni}) + \langle \psi_j | \nabla_n^2 | \psi_i \rangle \chi_{ni} \right\} + \nabla_n^2 \chi_{nj} + \epsilon_j \chi_{nj} = E_T \chi_{nj}$$

The terms in the curly bracket are called the first- and second-order non-adiabatic coupling elements. They represent the coupling between different electronic states and become important for those processes which involve more than one electronic state, e.g. in the case of conical intersections³.

Usually we use the adiabatic approximation, where only one electronic state is considered and all coupling terms are neglected. This means that only the $i = j$ terms survive. For a spatially non-degenerate wave function, the diagonal first-order non-adiabatic coupling element is zero[50]:

$$\langle \psi_i(r, R) | \nabla_n | \psi_i(r, R) \rangle = 0 \quad (2.7)$$

This allows us to rewrite the Eq.(2.1), for a given set of nuclear coordinates R , in the form:

$$(T_n + \epsilon_j(R) + U(R)) \chi_{ni} = E_T \chi_{ni} \quad \text{with} \quad U(R) = \langle \psi_i(r, R) | \nabla_n^2 | \psi_i(r, R) \rangle \quad (2.8)$$

In the *Born-Oppenheimer* approximation, we even neglect the *diagonal correction* term, $U(R)$, and as a result the motion of nuclei and electrons are fully decoupled. For every nuclear configuration R , we solve the electronic Schrödinger equation to obtain the electronic energy contribution, $\epsilon_j(R)$. The latter plays the role of a potential energy for the nuclear Schrödinger equation,

$$(T_n + \epsilon_j(R)) \chi(R) = E_T \chi(R) \quad (2.9)$$

²In bra-ket notation this can be written as:

$$\int d\mathbf{r} \psi_i^*(r, R) \psi_j(r, R) = \langle \psi_i(r, R) | \psi_j(r, R) \rangle = \delta_{ij} \quad (2.6)$$

³This occurs when two states of the same symmetry cross. A conical intersection seam is defined as the region, defined by a set of structures, where the potential energy surfaces of the two crossing states meet. This degeneracy can be lifted along two directions, namely the gradient difference vector and the gradient of the interstate coupling vector of the two states. Geometry optimization in the subspace spanned by all degrees of freedom except those mentioned ($\rightarrow n-2$, where n is the total number of degrees of freedom), leads to a 2-dimensional double-cone shaped PES. The point where the two cones intersect is thereafter a $n-2$ 'funnel' where two states are energetically degenerate. For more details see Refs.[51, 52] and references therein.

For a given electronic state, j , we allow the nuclei to move on a *potential energy surface* (PES). Thermally driven reactions for example occur on the electronic ground state PES, whereas in photochemical reactions both electronic ground and excited states are involved.

The advantage of using quantum mechanics (QM) over molecular mechanics (MM) (see below), is that those electronic processes, which are of high interest in chemistry and biology, such as bond formation, bond breaking, electronic excitations, charge transfer, energy transfer, etc, are essentially included in the quantum description of matter.

Furthermore any molecular property can be calculated by considering the first, second and higher order derivatives of the energy with respect to perturbations like external electric or magnetic field⁴. We are often interested in the derivative of the energy with respect to nuclear coordinates, usually referred to as *gradient*, which is used for locating minimum energy structures on the PES. These are stable molecular configurations and can be interpreted as reactant and possible product(s) of a chemical reaction. In addition, we can find saddle point(s) on the PES which correspond to the transition state structure(s) of such a chemical reaction.

Of special interest in the context of performing geometry optimization is the use of analytical gradients, which were first introduced by Pulay (see Ref. [53]). This method, which relies on the assumption that the PES is analytical (hence expandable in a Taylor series of nuclear displacements ΔR), is implemented in most of the quantum chemical codes and allows for very efficient geometry optimization procedures.

In the QM description, the level of accuracy at which molecular properties and chemical reactions/processes are described strongly depends on the wave function ansatz used to solve the electronic Schrödinger equation. For rather small molecules, the wave function ansatz can be systematically improved and we can afford calculations with *chemical accuracy*⁵. For a single conformation of a 'medium size' protein/DNA however, it is not yet feasible to solve the Schrödinger equation at a satisfactory level, let alone scanning the potential energy surface.

Molecular Mechanics (MM)

If we are only interested in the different conformations (e.g. possible orientations of amino acids in a given protein) and not in the electronic changes

⁴A nice overview is provided in Chapter 10 of Ref. [50]

⁵i.e. typical accuracy of a chemical experiment: 1 kcal/mol for relative energies and reaction barriers and 1% error for geometrical parameters

within our bio-system, we may treat all atoms with Molecular Mechanics (MM). Here, every atom is characterized by a set of parameters such as (partial) charge and van der Waals (vdW) radius. Electrons do not enter the MM treatment explicitly, which means that the last mentioned parameters have to be determined/optimized in such a way that the corresponding non-bonding electrostatic and van der Waals interactions are described adequately. Since the electrons are also responsible for bonding, additional parameters such as bond, angle, and torsion force constants are also necessary to account for the integrity and connectivity of the bio-system. We have replaced the (QM) wave function $\psi(r; R)$ with a set of (MM) force-field parameters, hence we must redefine the potential energy surface accordingly. The corresponding MM-PES is defined by the MM energy which reads:⁶

$$E_{\text{MM}} = \sum_{A,B} \left\{ \epsilon_{AB} \left[\left(\frac{\sigma_{AB}}{r_{AB}} \right)^6 - \left(\frac{\sigma_{AB}}{r_{AB}} \right)^{12} \right] + \frac{1}{4\pi\epsilon_0} \frac{q_A q_B}{r_{AB}} \right\} + \quad (2.10)$$

$$\sum_{\text{bonds}} k_b (d - d_0)^2 + \sum_{\text{angles}} k_\theta (\theta - \theta_0)^2 + \sum_{\text{torsions}} k_\phi [1 + \cos(n\phi + \delta)]$$

This energy expression allows us to compute the (MM) energy for large systems, say biological macromolecules like DNA or protein, at a very low computational cost. Having an initial structure of a protein, we can use the energy expression above to find the nearest *local* minimum on the MM-PES. Ideally one would like to find the energetically lowest minimum, i.e. the *global* minimum, but this is an impossible task for large systems like proteins. Force fields are therefore used to scan some parts of the PES and sample a set of local minima or an *ensemble* of structures. Properties of the system under study can be then extracted from ensemble averages⁷.

There are several techniques which can be used for sampling (see Ref. [55] and references therein). The choice of the method strongly depends on the type of force field being employed and the size of the system (or number of degrees of freedom involved). Systematic search methods, which allow

⁶For the terms in curly brackets, the summation runs over non-bonded pairs. σ_{AB} and ϵ_{AB} are the vdW parameters and r_{AB} is the non-bonded distance. q is the atomic partial charge. d_0, θ_0, ϕ_0 are values for the equilibrium bond distances, angles and torsions respectively. For a more detailed description of these terms and parameters see Ref.[54] and references therein.

⁷Ensembles are characterized by the parameters which have been kept fixed during the simulation. These include number of particles N , energy E , volume V , temperature T , pressure P and chemical potential μ . Depending on the combination of fixed parameters we can create Canonical, Micro-canonical, Isothermal-isobaric and Grand-canonical simulations.

for the complete scan of the PES, are only feasible for very small molecules. Instead one can use random search methods like Monte-Carlo (MC), where the occurrence of the conformation produced (randomly) by the algorithm, at temperature T , is proportional to its Boltzmann factor, $\exp(-\frac{\Delta E}{k_B T})$ [55].

Another way to sample the conformation space is to add dynamical effects to the system. This can be done by applying Newton's (\rightarrow molecular dynamics simulation, MD) or Langevin's (\rightarrow stochastic dynamics simulation, SD) equations to the motion of the atoms. By integrating these equations over time, we obtain a sequence of structures, the so called *trajectory*, which results from the propagation of the initial structure on the PES.

In MD (or SD) simulations the size of the integration time step has to be short enough to capture the vibration of the fastest particle in the system, e.g. hydrogen atoms. Therefore a time step of 1 fs is used in classical MD simulations. By fixing the hydrogen atoms, e.g. by using the SHAKE algorithm[56], it is nowadays possible to extend the time step and perform longer simulations. Although we are still unable to scan the whole conformation space, but if the duration of simulation, i.e. the number of integration steps, is long/high enough, it is possible to collect structures which are more or less representative for at least some part of conformation space⁸.

Quantum Mechanics/Molecular Mechanics (QM/MM)

The whole concept of the Quantum Mechanics / Molecular Mechanics (QM/MM) scheme, initiated by the work of Warshel and Levitt[58], is based on combining the advantages of both QM and MM treatments of matter. Within this approach we can afford to study relatively larger systems and at the same time have made a compromise between feasibility and accuracy[59, 60, 61, 62, 63].

The basic idea is to separate the system into a QM and a MM section. The QM part, or the inner region I (see Fig. 2.1), contains those atoms which will be handled with quantum mechanics (i.e. using a wave function). It is evident that this region should contain the active center of a chemical reaction. The second part, the outer region O, consists of those atoms which are to be treated classically (via predefined parameters). Since we may have to cut through covalent bonds in defining a QM region, we

⁸A further method is the simulated annealing[57]. Here we start a MD (or MC) simulation at very high temperatures which allows us to scan all of the conformation space. The temperature is gradually reduced, hence trapping the structure in a minimum. In order to find the global minimum, one would have to reduce the temperature infinitely slowly. This on the other hand implies an infinitely long simulation which is practically impossible.

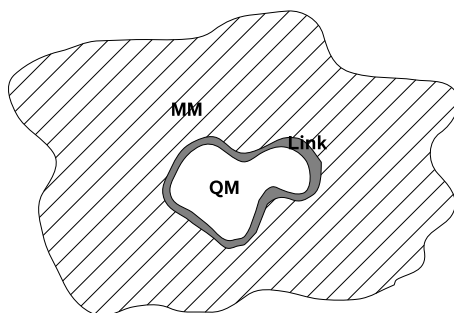


Figure 2.1: Schematic view of a large molecular system which can be divided in to QM, MM and Link regions.

usually introduce a third region, the so called link region L. This is used to cap the broken bonds and its treatment strongly depends on the QM/MM scheme being employed.

There are two main QM/MM schemes, namely *additive* and *subtractive*, which differ in the way that the energy expression for the total energy of the system, $E_{QM/MM}^T$, is defined. In the subtractive scheme, the total QM/MM energy is defined as

$$E_{QM/MM}^T = E_{MM}^{I+O} - E_{MM}^{I+L} + E_{QM}^{I+L} \quad (2.11)$$

In order to compute this energy, we first calculate the MM energy of the whole system (without any link region!). The MM and QM energies of the inner and link regions are then calculated respectively and plugged into the Eq.(2.11). The method is rather simple to implement as there are no modifications to the QM and MM codes required and there is no explicit QM-MM coupling term to be calculated⁹. A disadvantage of the subtractive scheme is that the non-bonding interactions between the inner and the outer region are treated at the MM level (first term of Eq.(2.11)). For example the electrostatic interactions, which change dramatically during a chemical reaction (for both thermal and photochemical reactions), can not be described by the fixed partial charges of the inner and outer atoms. Furthermore the QM charge density is not polarized by the MM/outer region¹⁰. More sophisticated subtractive schemes such as IMOMO[64] or ONIOM[65], have been suggested by Morokuma and coworkers. These schemes include additional QM-MM coupling terms and partially solve the last mentioned problems.

⁹QM-MM coupling terms represent the interaction between the atoms in the QM and MM regions.

¹⁰The outer region does not enter the third term of Eq.(2.11).

In the additive QM/MM scheme the total QM/MM energy of the entire system is written in the form

$$E_{QM/MM}^T = E_{MM}^O + E_{QM}^{I+L} + E_{QM-MM}^{I,O} \quad (2.12)$$

Here, E_{MM}^O is the MM energy of the outer region and E_{QM}^{I+L} , the QM energy of the inner and link regions. The third term, $E_{QM-MM}^{I,O}$, is the QM-MM coupling term and represents the interactions between the inner and outer atoms¹¹. These include QM-MM electrostatic, QM-MM van der Waals and QM-MM bonding terms.

The electrostatic coupling can be defined at different levels of theory. In a *mechanical* embedding we treat the electrostatic interactions at the MM level which corresponds to the subtractive scheme. In an *electrostatic* embedding, the QM energy is calculated in the presence of the (MM) point charges which enter the Hamiltonian as one-electron terms:

$$\hat{H}_{QM-MM}^{el} = - \sum_{i \in QM} \sum_{A \in MM} \frac{q_A}{|r_i - R_A|} + \sum_{\alpha \in QM} \sum_{A \in MM} \frac{q_A Z_\alpha}{|R_\alpha - R_M|} \quad (2.13)$$

where q_A and R_A are the charge and position of the MM atoms. r_i is the electron coordinate and the summation runs over all electrons. Z_α and R_α the nuclear charge and position of a QM atom.

The advantage of such a scheme is that the QM region is polarized by the MM region¹². Furthermore the MM energy of the inner region will not be calculated, hence no force field parameters for the atoms of the inner region are required.

Another option is to use a *polarized* embedding which involves polarizable force fields. Here we employ flexible MM point charges which can be influenced/polarized by the QM charge density[59].

The QM-MM van der Waals interactions, due to their short-range nature, are not as problematic as the electrostatic ones. Although every inner atom is in vdW interaction with all outer atoms, but only the close inner-outer pairs exhibit significant energy contribution. The Lennard-Jones potential, as seen Eq.(2.10), is typically used to calculate the QM-MM vdW coupling which corresponds to a pure MM treatment of this interaction. One should say that we are still facing the problem of changing vdW interactions during a chemical reaction. But the errors caused in such treatment can be

¹¹Note that here, in contrast to the subtractive scheme, the QM-MM coupling is explicitly included in the energy expression.

¹²This leads to over-polarization of the QM region of MM boundary atoms if we were to cut through a covalent bond. Possible solutions are discussed in Ref.[59].

reduced by expanding the QM region and thereafter lowering the number of 'sensitive' QM atoms which are in vdW interaction with the outer region atoms. Possible ways of introducing a link region and the treatment of its electrostatic/vdW/bonding interactions with the QM region is illustrated in Ref. [59] and will not be discussed here. In addition, a valuable survey of biomolecular QM/MM studies can be found in Ref. [59].

A detailed coverage of the different aspects in quantum chemical calculations, molecular dynamics simulations and QM/MM calculations has been avoided here as there are enough good reviews and books on these topics. The remainder of this chapter is meant to provide a short tutorial, based on personal experience gained as part of the PhD project, which might help beginners set up and perform similar QM/MM calculations.

2.2 Preparation of QM/MM input structures

It is recommended to use an experimentally determined structure as a starting point for the QM/MM calculations. The main reason is that a global optimization of the structure of a biomolecular system such as DNA or protein is very a demanding task¹³. One has to bear in mind that an experimentally determined structure represents only one possible conformation which may have been strongly influenced/dominated by the conditions imposed by the experiment (e.g. pH, temperature, pressure, chemical additives, etc). In other words, we may regard this structure as a single point on the PES under certain experimental conditions¹⁴.

In order to probe other regions of the conformation space, with the hope of finding other stable conformations/local minima, barriers need to be overcome. These barriers can be surmounted by introducing and consequently raising the kinetic energy of the system during the simulation. The ensemble of structures obtained from a simulation can be used to select input structures for the QM/MM calculations.

In this project, we have performed MD simulations at constant temperature and volume conditions (\rightarrow NVT) which yield canonical ensembles. In the following, steps taken to create such an ensemble, as employed to study

¹³Having a sequence of amino acids, one can try to follow the folding pathway of the corresponding protein by performing molecular dynamics simulations. We can assume that the process of folding occurs within $1\mu\text{s}$ - 0.1 s [55]. The number of MD steps, assuming that we know the folding pathway in advance!!, is in the order of 10^9 - 10^{14} and requires enormous computer resources and months of simulation.

¹⁴to be more exact it is a local minimum in which the system has been trapped.

BLUF domain proteins are outlined. The MM-code CHARMM[66, 67, 68, 69, 70] was used in this study since an MM-interface for CHARMM parameters, DL_POLY[71], is already implemented in the QM/MM interface ChemShell[72]. Helpful commands used in the CHARMM and ChemShell packages are printed in TYPEWRITER and SANS-SERIF formats, respectively.

1. **Initial structure:** If the initial geometry of the protein (or DNA) is to be extracted from an x-ray diffraction experiment, hydrogen atoms are not resolved and these have to be added manually. MM codes like CHARMM have built-in modules for this purpose(HBUILD).

An important issue to mention at this stage is the **protonation state** of amino acids when working with crystal structures of proteins. Depending on the quality of the experimentally resolved structure, one may have to change the protonation state of certain amino acids¹⁵. The change of protonation can be carried out using the built-in modules of the corresponding MM codes (PATCH).

A further aspect to keep in mind is that the whole system should preferably be kept electrostatically neutral¹⁶. This avoids 'dragging' an unnecessary total non-zero charge in the MM part of the QM/MM calculations, which is problematic if the QM region and the amino acid with the 'unbalanced' charge happen to be too close to each other.

2. **Solvent constraint:** Solvent molecules (in most cases water) should be added to the structure. A box or sphere (depending on the boundary conditions which are to be imposed), containing solvent molecules are superimposed on the initial bio structure¹⁷. Those water molecules which are either strongly overlapping with, or too close to the protein atoms, are removed. In the case of a DNA chain, counter ions (K⁺ or Na⁺) should also be added. If a water box is used for solvation, periodic boundary conditions should be enforced, so that no water molecule is allowed to 'escape' the system during the

¹⁵If there are no clear hints for a different protonation state than stated in the published structure, one can also consider all possible H-bonding networks and minimizing the energy of their corresponding structures. The protonation state in the most stable conformation should be then used in the simulations. Additionally, programs like PROPKA (<http://propka.ki.ku.dk/drogers>) or WHATIF (<http://swift.cmbi.kun.nl/whatif>) can be used to predict the pK_A values.

¹⁶This can be done by appropriate protonation of residues in the close neighborhood of the charged amino acids.

¹⁷For this purpose, the water clusters in the box or sphere should be pre-optimized.

simulation. As for the water sphere which was employed in this study, a harmonic potential is imposed on all atoms on the surface of the sphere¹⁸. Note that this constraint should be applied *throughout* the MM calculations and it may only be lifted at the QM/MM stage where the MM atoms, specially those at the system boundary which should ideally be far from the QM region/atoms, are not expected to change their positions dramatically.

3. **Force Field parameters:** After adding the complementary atoms (hydrogen, cations, solvent molecules, etc), force field parameters have to be assigned. Special care has to be taken when dealing with substrate-protein interaction, which is a typical case studied by QM/MM calculations. Although force field parameters are available for amino or nucleic acids[66, 67], this is not always the case for the substrate. There are two possible solutions to the problem.
 - (i) The parameters may have to be determined. Here it is absolutely crucial to obtain the parameters in the same manner as they were determined for the amino or nucleic acids¹⁹.
 - (ii) The parameters can also be 'reconstructed' from those available by analogy. For a given atom in the substrate, we consider its bonding partners or its connectivity (and possibly its charge). Having the 'chemical environment' of the chosen substrate atom in mind, we shall then have to find an atom type from the list of atom types already defined by the force field (e.g. parameters/atom types assigned to amino acids or nucleic acids), which matches our desired specifications. As an example, the parameters used for the isoalloxazine part of the flavin-mono-nucleotide in the QM/MM study of BLUF domains, is presented here (Fig 2.2 and Tab. 2.1).
4. **Initial geometry optimization (MM level):** Once the structure and the parameters are available, an initial geometry optimization should be carried out. Hereby the biomolecule-solvent complex is allowed to find its nearest minimum energy conformation. This step removes any bad contacts within the biomolecule induced by the experiment (e.g. crystallization). In addition any unfavorable interaction between the solvent molecule (and perhaps counter ions when studying DNA) and the biomolecule can be removed. In the case of BLUF

¹⁸The MMFP option of CHARMM is rather useful for this purpose.

¹⁹Detailed information how the CHARMM force field parameters are obtained can be found in Ref.[66].

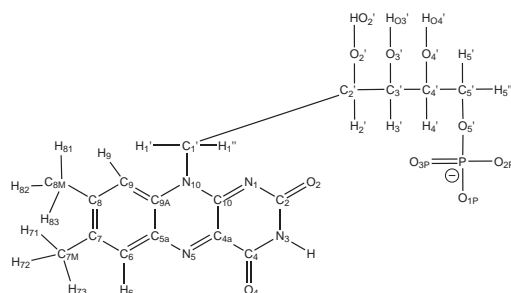


Figure 2.2: The atom names of the FAD chromophore.

Name	Type	δq	Res.	Name	Type	δq	Res.
N1	NN3	-0.66	CYT	C1'	CN8	-0.18	Rib
C2	CN1	0.52	CYT	H1'	HN8	0.09	Rib
O2	ON1	-0.49	CYT	H1''	HN8	0.09	Rib
N3	NN2	-0.46	URA	C2'	CN7	0.14	Rib
H3	HN2	0.36	URA	H2'	HN7	0.09	Rib
C4	CN1	0.53	URA	O2'	OH1	-0.66	Rib
O4	ON1	-0.48	URA	HO2'	H	0.43	Rib
C4A	CN2	0.36	GUA	C3'	CN7	0.14	Rib
N5	NN3	-0.66	GUA	H3'	HN7	0.09	Rib
C5A	CN2	0.36	GUA	O3'	OH1	-0.66	Rib
C6	CN3	-0.115	PHE	HO3'	H	0.43	Rib
H6	HN3	0.115	PHE	C4'	CN7	0.14	Rib
C7	CN3	0.00	PHE	H4'	HN7	0.09	Rib
C7M	CN9	-0.27	ALA	O4'	OH1	-0.66	Rib
H71	HN9	0.09	ALA	HO4'	H	0.43	Rib
H72	HN9	0.09	ALA	C5'	CN8	-0.08	Rib
H73	HN9	0.09	ALA	H5'	HN8	0.09	Rib
C8	CN3	0.00	PHE	H5''	HN8	0.09	Rib
C8M	CN9	-0.27	ALA	O5'	ON2	-0.62	Rib
H81	HN9	0.09	ALA	P	P	1.50	Rib
H82	HN9	0.09	ALA	O1P	ON3	-0.68	Rib
H83	HN9	0.09	ALA	O2P	ON4	-0.82	Rib
C9	CN3	-0.115	PHE	O3P	ON3	-0.82	Rib
H9	HN3	0.115	PHE	H3T	HN4	0.34	Rib
C9A	CN2	0.36	GUA				
N10	NN2	-0.05	ADE				
C10	CN2	0.31	GUA				

Table 2.1: The CHARMM parameters constructed for the (FMN) where the atom names, types and partial charges δq are given. The origins of the assigned types and charges are also given (abbreviated names for the nucleic and amino acids, GUA = Guanine, ALA = Alanine, Rib = ribityl side chain in nucleic acids, etc.). Usually, for a given atom type, the choice of partial charge is unique. For the special case of FAD however we had to modify some of the partial charges slightly (e.g. N5-atom). The reason is that all the partial charges defined for a residue must add up to the total charge assigned to that residue and the default partial charges do not fulfill this condition automatically. This modification should be rather safe, as long as the atoms with such modified partial charges are not in H-bonding interaction (for QM/MM calculations, where the flavin and its closest neighbors are treated quantum mechanically, this change should not make much difference).

domains, this initial geometry optimization was separated into different stages:

- (i) All atoms except hydrogens were kept fixed. This allows for the adaptation of these lightest particles to the heavier atoms to which they are attached during the protonation step.
- (ii) The substrate, here being flavin, was then allowed to relax its structure to the structure of the protein.
- (iii) Now the geometry of the whole system can be relaxed. Optionally, the protein backbone atoms can be constrained and then allowed to relax in second step.

5. **Heating and equilibration (MM level):** As mentioned earlier, in order to find other conformations it is necessary to perturb the system from its initial conditions and allow it to propagate in the conformation space by assigning momenta/velocities to all atoms. The kinetic energy of the system, E_{kin} , is distributed among all atoms and is related to their assigned velocities $v_i(t)$ via the relation

$$E_{kin}(t) = \sum_i \frac{1}{2} m_i v_i^2(t) = \frac{1}{2} N k_B T(t) \quad (2.14)$$

This relation can be used to define the temperature, T , for a system with N degrees of freedom. This temperature can be raised/lowered at any time during the simulation by rescaling the velocities of the particles accordingly. To begin with we can set the temperature to 50K (FIRSTT) and raise it gradually (TEMINC) at certain time intervals (IHTFRQ) until the final (room) temperature is reached (FINALT).

The backbone of protein/DNA should not be allowed to change its conformation drastically during the course of heating as this may lead to unrealistic structures. In order to avoid such effects, constraints can be applied to the backbone atoms by keeping them in a harmonic potential with a high force constant (CONS HARM). The backbone constraints may be gradually lifted once the temperature has reached the desired level. At this last stage, the structure has been adapted to the room temperature conditions (equilibration) before we start with the actual ensemble production MD-stage.

There is an important issue which has to be addressed before going to the next step. There is a high probability that the water box/sphere which has been used to solvate the biomolecule is deformed during the equilibration run. Cavities can be created by the backbone and/or the residues as a result of the conformational changes. If these cavities are of hydrophilic nature (i.e. with high hydrogen bonding

capability), then water molecules fill these empty spaces, leading to the deformation of the solvent box/sphere. It is therefore recommended to re-solvate the biomolecule by superimposing the present structure with the initial water box/sphere and remove the overlapping water molecules. One should then repeat the whole procedure from the heating/equilibration stage. This cycle should be repeated until no (or at least only few) water molecules are necessary for the re-solvation process.

6. **MD Simulation/Production Run:** In the production stage of the MD simulation we allow the system to evolve with time. For a canonical ensemble we need to keep the volume and the temperature constant. The volume is controlled by the harmonic potential imposed on the water molecules on the border of the solvent sphere as described previously. In addition a thermal bath is introduced in the simulation which can regulate the temperature via thermal coupling (TCOUpling)²⁰. The allowed deviation from FINALT can be specified using the TWINDH and TWINDL commands.

The issue usually raised here is the duration of the MD production run. Of course the longer the simulation, the higher the chance of sampling structures which are uncorrelated with those already present in the trajectory. In QM/MM calculations, we are usually interested in the chemical processes which occur in the inner/QM region. We therefore prefer to sample structures which either differ in the conformation of the inner region or in the interaction between the inner and outer regions. Finally we have to pick out structures from the trajectory which are then used for the QM/MM input calculations²¹

7. **Final minimization (MM level):** After choosing a set of structures from the MD run, a further and final geometry optimization, at the force field level of theory, is necessary. The reason is that these structures are only snapshots taken from the trajectory and do not represent a minimum energy geometry.

Furthermore huge savings can be made on the CPU time during the QM/MM geometry optimization. We assume that those MM atoms far enough from the QM region are not affected by the QM region

²⁰For example the excess energy, leading to temperatures higher than FINALT, is absorbed by the bath.)

²¹There are several ways for choosing the structures. These include systematic stratified sampling, random stratified sampling and coarse grained sampling[50].

in the QM/MM calculations and are therefore subject to pure MM interactions. We can further assume that a rather large number of MM energy and gradient steps are needed (say 2000) to optimize the geometry of this specific part of total system. If the structure obtained from the MD run is used directly in the QM/MM calculation, one would have to carry out at least the same number of QM/MM energy and gradient steps. Considering that the QM step is the most expensive part of a QM/MM calculation, it is therefore not desired to perform 2000 QM/MM steps which can otherwise be avoided by just pre-optimizing the structure at the MM level of theory.

2.3 ChemShell interface

The ChemShell package[72, 73] provides an environment for the communication between different QM and MM codes during a QM/MM calculation (see Fig. 2.3). Based on the Tcl interpreter[74], it is possible to transfer or translate (changing the format of the output/input data) the necessary data such as coordinates, energy values, QM- and MM-gradient vectors, point charge field and force field parameters from one code to another. It also contains helpful built-in modules which can be used to setup and perform QM/MM calculations. These include adding link atoms (capping hydrogen), calculation of QM-MM coupling terms, enforcing constraints on bond length, bond angle, dihedral angles, etc. One of the most important features is the ability to perform geometry optimizations.

In this project, the ChemShell interface was used to carry out single point QM/MM energy calculations and QM/MM geometry optimization with and without restraints. Slight modifications were made to the TURBOMOLE-interface such that geometry optimizations of excited states at the TD-DFT level of theory could be carried out²².

2.4 QM/MM Geometry optimization setup

ChemShell provides several optimizer modules, which depending on the size of the system and type of structure search (e.g. minimum energy or saddle points) can be used. For large systems with expensive QM-steps, the HDLCO-pt module is recommended. For this purpose a set of residues (not to be mixed with the usual definition of a residue in a protein/DNA)

²²the ChemShell script was mainly forced to read excited state energy and gradient instead of the corresponding values of the ground state.

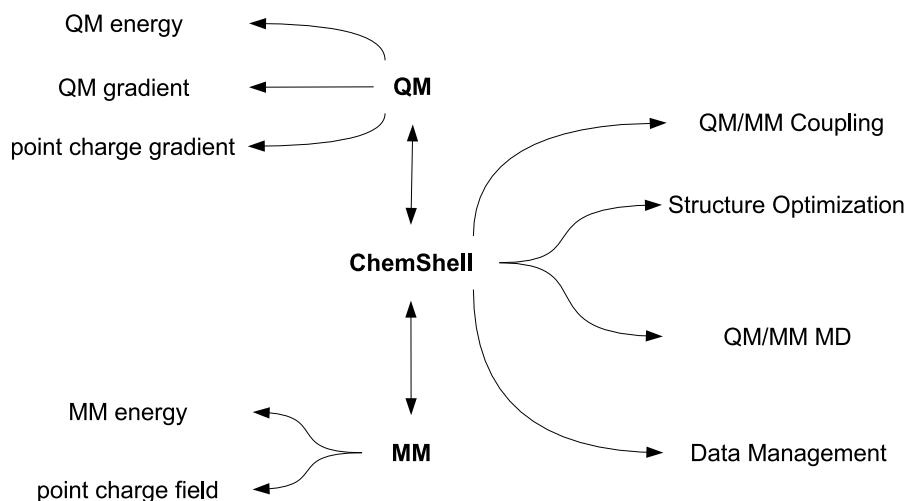


Figure 2.3: Communication between different parts of QM and MM calculations provided by ChemShell. QM and MM codes for which an interface is currently available include: GAMES-UK, MNDO04, MOPAC, TURBOMOLE, Gaussian, MOLPRO, Orca, NWChem, DL-POLY, CHARMM, GROMOS and GULP. For more information see www.chemshell.org

is defined by reading the Cartesian coordinates of the system²³. Within each **residue**, delocalized internal coordinates[75] are defined and used for the optimization procedure. ChemShell also carries out the transformation between Cartesian and delocalized internal coordinates.

In practice, it is not necessary to optimize the geometry of the whole system and one may define a region (containing QM, Link and MM atoms) around the chemically active part for which the geometry optimization should be carried out. In the case of BLUF domains for example, we took all protein residues and water molecules which lie within a sphere of about 25Å around the flavin chromophore. In order to declare those atoms which should be optimized, the keyword **active_atoms** is used. For atoms which should be kept fixed, the keyword **frozen** is used instead. The QM region (i.e. list of QM atoms) is defined by the keyword **qm_region** (for the choice

²³once the keyword **residues** is set, ChemShell will use the connectivity and define the required residues automatically

of QM region and termination, see further below).

The convergence criteria can be set manually (see ChemShell manual), but the default settings (HDLC: Gradient-RMS $0.3E-03$, Cartesian: Gradient-RMS $0.3E-03$) is sufficient in most cases.

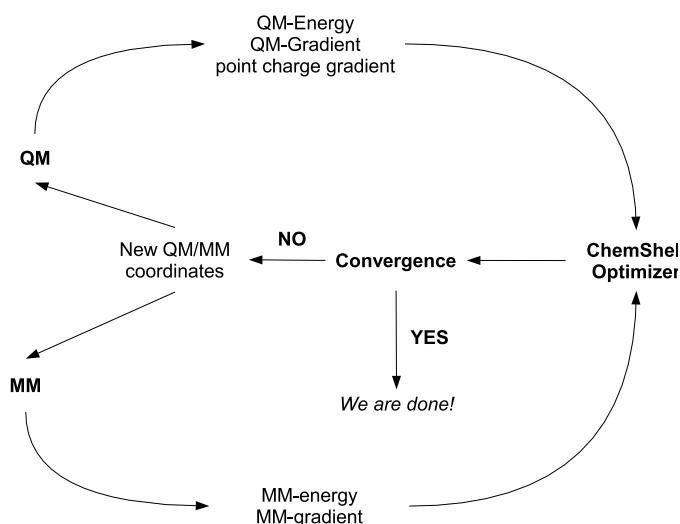


Figure 2.4: The cycle used in QM/MM geometry optimization calculations.

2.5 Choice of QM region

The choice of QM atoms (\rightarrow QM region) is a further issue which has to be dealt with great care. Those atoms being directly involved in a chemical reaction, hence undergoing significant changes in their electronic state, should be the main part of a QM region. For example if an electronic excitation of charge transfer character is studied, both electron-donor and electron-acceptor molecules/residues should be included in the QM region. These changes will certainly have influence on the neighboring atoms, which should ideally be described at the QM level too. One may also consider adding those residues to the QM region which interact with the chemical region of interest via hydrogen bonding. In the case of strong electrostatic interactions, e.g. between counter ion and ribose in the case of DNA, it is recommended to include both sides in the QM region. Another important aspect is the termination of a QM region, or in simple words the question of how and where to cut a QM region. First of all

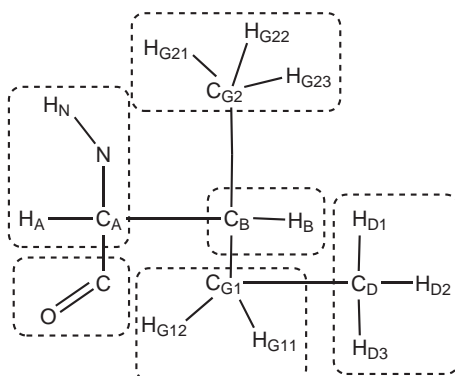


Figure 2.5: Point charges defined for the ILE residue in CHARMM. The residue is divided into different groups (shown in squares). In each group the point charges of the atoms are chosen such that the net charge is a rational number. Total charge of the residue is the sum of all these net charges.

one should avoid cutting through a polar (e.g. C-O or C-N) bond, or an aromatic system (e.g. benzene ring), instead ‘innocent’ unpolar C-C bond should be broken. One should also verify that no net charge is being created on the MM side. In CHARMM for example the point charges of the atoms in each residues is distributed in small groups of net zero charge (Fig. 2.5) and it is recommended not to disturb these charge groups. The dangling bonds in the QM region should now be capped. A popular approach is to use the so called *link atoms* which are mostly hydrogen[76].

2.6 Choice of QM theory: Ground state

In principle there is a wide range of methods for describing the electronic ground state and they are usually categorized into semi-empirical methods, electron-density based methods (mainly Density Functional Theory/DFT) and wave function based methods. A nice overview of all these methods can be found in Ref. [50].

The main challenge is to provide a correct description of non-bonding interactions which govern biologically relevant system such as DNA, protein and even water. These electrostatic (e.g. ionic, hydrogen bonding, induced electrostatic) and dispersive (stacking interaction) interactions however are not fully covered by all of the above mentioned QM methods.

Density functional theory[77], due to its much lower computational cost compared with wave function based approaches (post-HF [78]), has become a very popular method for both gas-phase QM and QM/MM ground

state calculations. Unfortunately the mostly used functionals of DFT (mainly B3LYP) are known to have problems describing weak/non-bonded interactions (see Ref.[79] and references therein). Indeed the performance of DFT is not so bad when it comes to predicting hydrogen bonding strength (\rightarrow H-bond length). But this in fact is due to a cancellation of errors. The latter however is not capable of producing tolerable results for long range van der Waals dispersive interactions²⁴. Attempts made to cure this problem of DFT has been discussed in a recent review by Černý and Hobza[83].

An alternative way would be to use pure *ab initio* methods[78] such as Møller-Plesset perturbation theory (MP) and Coupled Cluster (CC) theory, which show improvements with respect to Hartree-Fock (HF) ansatz in all aspects as they include systematic electron correlation²⁵. Singles and Doubles coupled cluster theory with perturbative treatment of triple excitations (CCSD(T)), if extended basis sets are employed, is a sufficiently accurate way to compute interaction energies and hence suitable for benchmarking purposes[88]. Møller-Plesset perturbation theory of second order (MP2), the cheapest *ab initio* method which includes electron correlation, tends to overestimate dispersion energy. An empirically corrected form of MP2 has been proposed by S. Grimme, where the correlation energy contribution of an electron pair with parallel spin scales differently from that of a pair with anti-parallel spin[89]. The so called spin-component-scaled MP2 (SCS-MP2) method, which is no longer an *ab initio* method, yields results which are in most cases²⁶ comparable with those obtained from the much more expensive CCSD(T) method[91].

In contrast to DFT methods where the computational cost of energy and gradient calculations is not a big issue, we usually face rather time consuming calculation steps in the case of high level wave-function based methods. Of course this is mainly due to the fact that post-HF wave functions include

²⁴There are new functionals developed (see for example Ref. [80]) which are meant to cure this problem. Grimme suggested to combine the DFT description with an explicit R^{-6} term[81] which describes the interatomic dispersive interaction. Morgado *et al.* have compared the results obtained with DFT-D with those from high level *ab initio* calculations[82].

²⁵These methods are all valid for single reference cases, i.e. no biradical, charge transfer or transition state is involved. Otherwise multi-reference variation of these methods, like CASSCF, CASPT2, MR-CI or MR-CC, should be used[84, 85, 86, 87].

²⁶A recent study by Bachorz *et al.*[90] has shown that the SCS-MP2 approach is not suitable for cases where the hydrogen bonding and $\pi - \pi$ stacking are competing against each other. SCS-MP2 reduces the interaction energy of both stacked and hydrogen bonded dimers. This is an improvement for the stacked dimer, but not for the hydrogen-bonded dimer

more than just one Slater determinant (\rightarrow many-electron-configurations). Moreover, if the canonical, and thus delocalized, HF molecular-orbitals are used to span the configuration space, we end up with an unfavorable scaling behavior of these methods with respect to the number of atoms present in the system.

One of the promising solutions to this problem is to use local (HF) molecular orbitals (MOs), obtained by a unitary transformation of the conventional/standard canonical MOs[92]. The argument here is that dynamic correlation effects, missing in the HF approximation, are short-range effects (for non-metallic systems) and the computational cost for treating them should therefore grow asymptotically only linearly with the size of the system[93]. By exploiting this feature, one can now perform local MP2 (LMP2)[94], CCSD (LCCD)[95] and CCSD(T) (LCCSD(T))[96], single point energy calculations²⁷ for which the computational effort grows linearly with the size of the molecular system. Another advantage of using local correlation methods is that they are almost free of the basis set superposition error (BSSE) and as a result no counterpoise correction to eliminate the error is required[98].

Although geometry optimization at the CCSD or CCSD(T) (and their local versions) is not yet possible, efficient MP2 and Local-MP2 gradients are already implemented in some of the currently available QM-codes. Perhaps the best choice for geometry optimization is the DF-SCS-LMP2 method[99], which has the advantages of density fitting²⁸, local and SCS approaches[88].

2.7 Choice of QM theory: Excited state

There are several *ab initio* approaches available for treating excited states. Due to size and number of the molecules which play a role in a biochemical process however, most of the accurate methods are not applicable. In this project two different methods, namely *Time-Dependent Density Functional Theory* (TD-DFT)[103, 104] and *time dependent second-order approximate coupled cluster singles and doubles model* (TD-CC2)[105, 106], were used to perform excited state calculations.

In both of these methods, linear response theory is used to obtain excitation

²⁷see Ref. [97] for a case study, where the barrier of an enzymatic reaction was determined with local method.

²⁸A further saving on CPU time (mainly for calculating repulsive two electron integrals) can be made with the help of the density fitting (DF) or resolution of identity (RI) approximation. For further details see Refs. [100, 101, 102] and references therein.

energies. Response methods in general can be used to obtain molecular properties of a system under static or time-dependent perturbations. The advantage of such a formalism is that the excitation energies and the corresponding properties of the excited states are obtained indirectly as a property of the electronic ground state and no direct reference to the excited state (and its wave function) is needed. For example it can be shown that electronic excitation energies are the poles of the frequency dependent ground state polarizability, which is the linear response of the dipole moment of the system subject to external electric field. The oscillator strengths of the corresponding excitations are the residues of the frequency dependent polarizability.

Although TD-DFT is computationally much more favorable than TD-CC2, it has its limitations. One of the common processes encountered in photobiology is the photo-induced charge transfer (CT) and it is well documented that TD-DFT is not reliable in this respect (see Ref. [107] and references therein). Compared with the valence excited states, the vertical excitation energies obtained for CT states are underestimated. This is caused by the self-interaction error in DFT, which becomes dominant if local-xc functionals (LDA, BP86, BLYP) are used. By adding fractions of exact HF-exchange [108] (\rightarrow B3LYP or BHLYP functionals), it is possible to improve the CT excitation energies. The TD-DFT optimized geometries of the CT excited states are less affected by the self-interaction error [109] and therefore reliable.

The TD-CC2 approach, based on linear response coupled cluster (LR-CC) theory [110], delivers vertical excitation energies which are typically correct within 0.3 eV [111, 106]²⁹.

The strategy which we have employed throughout this project is now briefly described. We are mostly interested in the reaction path on the charge transfer excited state³⁰. This reaction path can be defined by n constraints (e.g. fixed values for the internal coordinates such as bonds/stretch, angles/bends, dihedral angles/torsions and their linear combination(s)), which are imposed on the structure of the molecule during the course of the geometry optimization. By relaxing the geometry in all directions, except those defined by the constraints, we reduce the motion of our molecule

²⁹This is only true for excited state which are dominated by single excitations from the reference determinant [112]. A recent work by Schreiber *et. al.* compares the CASPT2, CC2, CCSD and CC3 methods for vertical excitations based on benchmark calculations carried out on a set of medium sized organic molecules [113].

³⁰In flavin based photoreceptors studied here, we have shown that a conical intersection between the locally excited state of flavin and the charge transfer state, is responsible for the activation of the latter [114, 45]. See chapters 3 and 4 for more details.

on the multi-dimensional PES down to a (lower) n -dimensional surface. The cheap TD-DFT method, due to its rather reliable excited state geometries (even for the CT state as mentioned above), can be used for this purpose. The reaction path obtained here, although having the a more or less "correct topology", is energetically underestimated. In order account for this error, we perform vertical excitation calculation at the TD-CC2 level for every TD-DFT relaxed structure on the reaction path. We this approach, we can identify structures on the CC2-based CT potential energy surface without having to perform excited state geometry optimization at the rather expensive CC2 level.

Chapter 3

Phenothiazine-Flavin Dyad

3.1 Introduction

With the aim of mimicking the function of a blue-light photoreceptor of cryptochrome type, a molecular arrangement of covalently linked pyrene, isoalloxazine and phenothiazine dye units (see Fig. 1.4(b)) was synthesised by Daub and coworkers[22]. Pyrene acts as an antenna, isoalloxazine (flavin) as an electron acceptor/redox mediator and phenothiazine is the electron donor of the system. The role of phenothiazine as an electron donor has been studied before[115]. Flavin-related dyes have been under investigation in the context of redox-enzyme mimetic[116, 117] or modelling natural photosynthesis[118]. In general similar types of molecular systems have been looked at over the years with the aim of improving our understanding of the electron transfer processes as well as developing molecular devices such as optical switches[119, 120, 121, 122, 123] or components for organic emitting diodes (OLEDs)[115, 124].

In the present work we are interested in the phenothiazine-phenyl-isoalloxazine dyad which is the donor-acceptor part of the proposed arrangement (see compound 3 in Ref[22]). Spectroelectrochemistry measurements carried out on this dyad[30] show the formation of a radical cation with the oxidation localised at phenothiazine and a radical anion with the reduction at the isoalloxazine subunit. Fluorescence measurements on this dyad show an extremely low fluorescence quantum yield ($\Phi_F = 0.0005$) upon excitation in the long wavelength band (463 nm = 2.68 eV), an indication for an efficient fluorescence quenching of the isoalloxazine moiety. In the time-resolved fluorescence measurements published by Schneider *et al.*[30] the short living component could not be resolved and only a long living component of 6 ns was reported.

More recent time-resolved measurements[24] determine a bi-exponential decay with a fast component of 600 fs and a slow component of 4.5 ns. A semi-empirical AM1 study on these dyads and triads has been reported before by Clark and coworkers [30, 31, 32]. Furthermore, the excited states of flavin and related molecules (uracil, isolumazine and lumiflavin) have been studied by Neiss *et.al* [125, 126] whereas for phenothiazine and its derivatives there are to our best knowledge only semi-empirical, DFT and HF calculations of the ground state available in the literature [127, 128, 129]. Here we present the results of an *ab initio* study on the excited states of the phenothiazine-phenyl-isoalloxazine dyad, which was slightly simplified for our calculations (*vide infra*). We have employed a combination of Time-Dependent Density Functional Theory (TD-DFT)[130, 131, 132, 109] on the one hand to optimise excited state geometries, and Coupled Cluster (CC2) Linear Response Theory[133, 134, 135, 105] to obtain reasonably accurate excitation energies. The underlying assumption is that equilibrium structures are much less affected than excitation energies by the notorious self-interaction problem of DFT for Charge Transfer (CT) states. Our calculations reveal a conical intersection between a Locally Excited (LE) and a (CT) state, which we assume to be the origin of the short lifetime and fluorescence quenching observed in the experiments mentioned above.

The fact that conical intersections[136, 137, 138] are important for the photochemistry of organic molecules is nothing new and has been investigated over the years in different systems in combination with various *ab initio* methods[139, 140, 141, 142, 143, 144, 145]. Conical intersections of excited states with the ground state provide highly efficient deactivation channel/funnels for the radiationless decay of these excited states. Nature thus exploits conical intersections of this type to stabilise vital biomolecules such as DNA against photo-degradation[146]. Theoretical and experimental investigations carried out on 2-aminopyridine dimers[147] mimicking Watson-Crick base pairs in DNA show a proton transfer step connecting the locally excited and charge transfer states with the ground state through two conical intersections[147, 148]. The short lifetime of the excited state (65 ± 10 ps) observed [147] in this system reveals the importance of such fast radiationless decay mechanisms which protect the living organism from the UV part of the sunlight. Similar situations are also encountered in studies on guanine-cytosine base-pairs[149, 150], 9H-Adenine isomer[151] and salicylic acid[152].

Furthermore conical intersections between two excited states may determine the photophysical behaviour of many molecules. One frequently discussed example of this type is 4-(Dimethylamino)benzonitrile (DMABN) which has been studied extensively both experimentally and theoretically

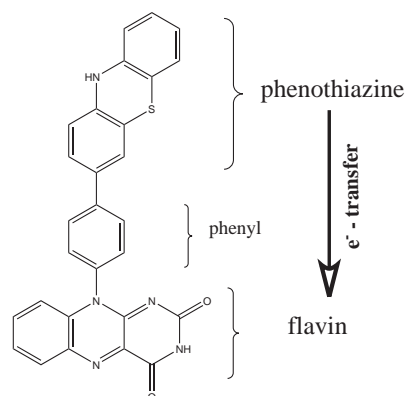


Figure 3.1: phenothiazine-phenyl-Isoalloxazine dyad.

(cf. Ref [153] and references therein). A CI between two excited states of different character, an LE and a CT state is responsible for the dual fluorescence of this molecule. A similar situation is encountered in the present study for the phenothiazine-phenyl-isoalloxazine dyad, though no fluorescence out of the CT state is observed here, due to the low transition strength of that state.

3.2 Computational details

The molecule investigated in the present study is depicted in Fig.3.1 and comprises 53 atoms (172 valence electrons), 36 of which are second row elements. Considering furthermore, that 153 degrees of freedom are to be relaxed in the geometry optimisations of ground and excited states, the system is quite large compared to other molecules which to our knowledge have been studied previously at the same level of theory. Hence a point to bare in mind at this stage is the limited feasibility of a detailed and accurate study on the excited state structures of this system. With the currently available codes we can only make qualitative predictions on the processes taking place. All TD-DFT and canonical CC2 calculations were carried out using the TURBOMOLE program suite[154]. Except stated explicitly otherwise the SVP basis set (split valence plus one set of polarisation functions) [155] together with the corresponding auxiliary basis [156] has been used. Ground and excited state geometries were optimised at the DFT and the TD-DFT level, respectively, using the BP functional[157]. Only minor changes (less than 1 pm) were observed in the equilibrium geometries on going to the B3-LYP[158] hybrid functional instead.

TD-DFT is of course very attractive due to its low computational cost (compared with correlated *ab initio* methods for excited states)[159], an important prerequisite for the treatment of extended molecular systems. However, it is well known that TD-DFT (with presently available local functionals) grossly underestimates excitation energies of CT states [148, 160, 161, 162, 163]. This error is directly related to the electronic self-interaction inherent in DFT [164, 165]. Since TD-DFT does not exhibit the physically correct $1/R$ asymptotic behaviour (R denotes the distance between the separated charges) the size of the error in the excitation energies depends strongly on R [161, 166]. For some cases (with relatively modest donor-acceptor separations) the use of a hybrid functional like B3-LYP, which partially brings back nonlocal exchange, alleviates this problem [167]. Yet in the present phenothiazine-phenyl-isoalloxazine dyad with a donor-acceptor separation of $R \approx 10.8\text{\AA}$, B3-LYP still delivers a qualitatively wrong physical picture with the CT excitation energies underestimated by more than 1.5 eV (*vide infra*). On the other hand, properties of excited states, such as minimum-energy geometries are much less affected by self-interaction [132, 109, 167].

An apparently reasonable compromise for studies on extended molecular systems such as the present one, would therefore be to combine TD-DFT geometries with correlated *ab initio* single-point energy calculations at these geometries. That is to compute *ab initio* potential energy surfaces along a certain subset of relevant coordinates, and relaxing the remaining degree of freedom at the level of TD-DFT.

In the present study we have employed Coupled Cluster Linear Response theory to calculate the required *ab initio* single-point excitation energies. Due to the size of the dyad only the simplest coupled cluster model, i.e., CC2 [105], could be afforded. In CC2 doubles substitutions are treated to first order in the fluctuation potential (just as in second order Møller-Plesset perturbation theory (MP2)) while singles substitutions are treated to all orders. CC2 is rather robust for excited states dominated by single excitations. It provides a qualitatively correct physical picture and the errors in the excitation energies usually do not exceed 0.3 eV. For the present study we employed the efficient implementation of Hättig and Weigend [168] available in TURBOMOLE, which is based on the Density-Fitting (DF) approximation for two-electron repulsion integrals (what they call the Resolution of the Identity (RI) approximation). For a recent comparison of TD-DFT and CC2 excitation energies we refer to Ref.[169]. Furthermore transition strengths and dipole moments of the excited states were computed at the CC2 level [170] for some points. The computational cost of the CC2 linear response method based on Density-Fitting scales as $O(N^5)$ with

the molecular size N in a canonical implementation. This makes calculations on molecules of this size or even bigger ones very costly, in particular for properties of excited states. Therefore, geometry optimisations at the level of CC2 are presently restricted to molecules of considerably smaller size [152, 171].

Presently, a new CC2 Linear Response method based on localised orbitals with a more favourable scaling of the computational cost is under development in our group. This Local CC2 (LCC2) response method [106, 172], implemented in the present development version of the MOLPRO package [173], has been used to study the effect of the solvent on the vertical excitation energies of the LE and CT states. For these local calculations the cc-pVDZ basis [174] together with the related fitting basis [175] was used. The pair list for the GS amplitudes included all pairs with interorbital distances not exceeding 10 Bohr, while the GS domains were determined according to a Boughton-Pulay criterion of 0.985. Pair lists and domains of the individual excited states, i.e., the related amplitudes response, were determined based on an analysis of an initial CIS (Configuration Interaction Singles) wavefunction, as described in detail in Ref.[106]. For this purpose the criterion for "important orbital" selection was set to $\kappa_e = 0.995$, while the related pair list included all pairs of important orbitals and pairs of all other orbitals up to an interorbital distance of 10 bohr. This is a rather conservative specification of the pair list, as shown in Ref.[106].

3.3 Results and Discussion

3.3.1 At the ground state geometry

The CC2 linear response calculations at the DFT optimised ground state equilibrium geometry reveal that the first excited state S_1 is of $\pi \rightarrow \pi^*$ type (HOMO-2 \rightarrow LUMO) and primarily localised on the isoalloxazine subunit (plots of the relevant orbitals are given in Fig. 3.2). It has the highest oscillator strength¹ ($f = 0.242$) within the five lowest calculated excited states of the dyad (cf. Table 3.1).

In the following we will synonymously use the notation *LE state* (locally excited state) for this S_1 state. The S_2 state again is primarily localised on the isoalloxazine subunit, however has an almost negligible oscillator strength. The S_3 state is the charge transfer state, dominated by the HOMO

¹The oscillator strength between states m and n is defined as $f_{mn} \propto \tilde{\nu}_{max} |\mu_{mn}|^2$, with $\mu_{mn} = \langle \Psi_m | \mu | \Psi_n \rangle$ denoting the corresponding transition dipole moment and $\tilde{\nu}_{max}$ the energy of the peak of the absorption band in wave numbers [33]

\rightarrow LUMO $\pi \rightarrow \pi^*$ orbital substitution with negative charge transferred from phenothiazine to isoalloxazine. It has consequently an appreciably higher dipole moment than the other states but its oscillator strength is again very small ($f = 0.01$).

Since the minimum-energy geometry of the CT state surface is markedly different from the ground state equilibrium geometry (*vide infra*) the related Franck-Condon factors are presumed to be small, as well. The next state has some CT character too, involving the phenothiazine and the bridging phenyl group (dominated by the HOMO \rightarrow LUMO+1 orbital substitution, see Fig.3.2), although its dipole moment calculated at the DFT/B3-LYP level is not significantly larger than that of the LE state. We will denote this state in the following as PCT (partial charge transfer) state. The fifth state finally is again localised on the isoalloxazine subunit and has $n \rightarrow \pi^*$ character.

Table 3.1: TD-DFT and CC2 vertical excitation energies (in eV, ordered according to the CC2 values), calculated at the DFT optimised ground state equilibrium geometry. Ground- and excited-state dipole moments (in Debye), as well as the related oscillator strengths (f) in length representation are also given.

		TD-DFT-BP		TD-DFT-B3LYP	
Character		E(eV)(f)	$ \vec{\mu}(\text{ex}) $	E(eV)(f)	$ \vec{\mu}(\text{ex}) $
S_0			9.136		8.973
LE	$\pi \rightarrow \pi^*$	2.683 (0.079)	11.669	3.029 (0.189)	13.018
LE	$\pi \rightarrow \pi^*$	2.457 (0.001)	16.185	3.154 (0.000)	15.126
CT	$\pi \rightarrow \pi^*$	0.953 (0.002)	45.466	1.834 (0.002)	46.034
PCT	$\pi \rightarrow \pi^*$	2.569 (0.131)	27.265	3.130 (0.154)	15.240
LE	$n \rightarrow \pi^*$	2.037 (0.001)	24.233	3.001 (0.001)	4.778

		CC2				
Character		E(eV)(f)	$ \vec{\mu}(\text{ex}) $	μ_x	μ_y	μ_z
S_0			8.366	6.81	4.63	-1.35
LE	$\pi \rightarrow \pi^*$	3.233 (0.242)	11.263	7.43	8.30	-1.62
LE	$\pi \rightarrow \pi^*$	3.416 (0.004)	15.286	5.16	14.25	-1.96
CT	$\pi \rightarrow \pi^*$	3.530 (0.010)	32.708	6.17	32.05	-2.09
PCT	$\pi \rightarrow \pi^*$	3.679 (0.202)	11.595	5.50	10.19	-0.54
LE	$n \rightarrow \pi^*$	3.701 (0.020)	10.082	0.48	9.86	-2.06

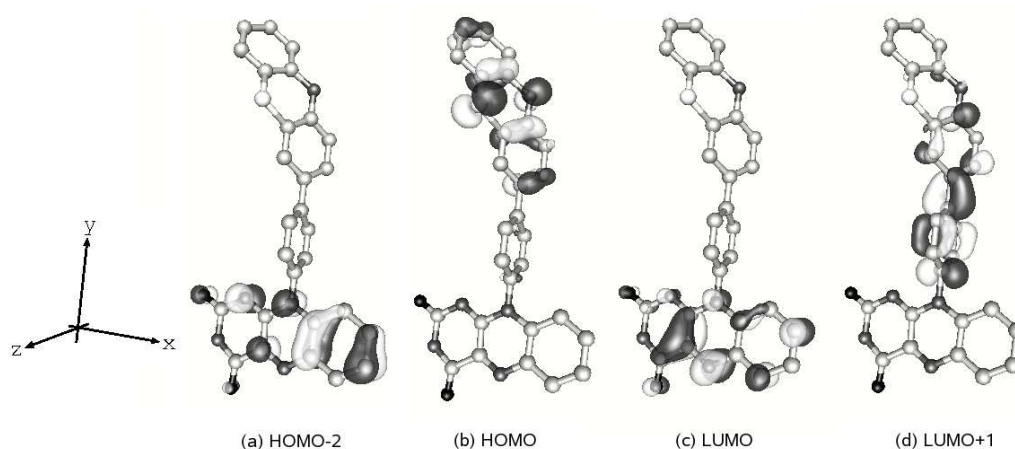


Figure 3.2: Hartree-Fock orbitals (calculated at the ground state geometry), which are relevant for the LE, CT, and PCT excitations (see text). The corresponding Kohn-Sham orbitals are very similar. The orientation of the molecule is such that the isoalloxazine subunit lies in the x-y plane.

3.3.2 At the CT minimum-energy geometry

The formation of the charge separated species (D^+A^-) observed in the experiments [30] leads us to the hypothesis that the "dark" CT state (with low oscillator strength) is responsible for the fluorescence quenching of the LE state. Since the CT state lies energetically above the LE state at the ground state equilibrium geometry and below the LE state at the CT minimum energy geometry, it can be anticipated that a conical intersection between these two states exists.

In contrast to the CC2 calculation TD-DFT provides an entirely different physical picture. Here, the "dark" CT state lies far below the "bright" LE state (by 1.2–1.7 eV). Depending on the functional the excitation energy for the CT state is 2.5 eV (BP) and 1.6 eV (B3-LYP) smaller than the corresponding CC2 value, clear indications of large self-interaction errors, also for the B3-LYP functional. For the PCT state the deviations between DFT and CC2 are a bit smaller, -1.11 eV and -0.35 eV for the BP and the B3-LYP functional, respectively, reflecting the shorter distance between donor and acceptor. The TD-DFT and CC2 dipole moments are quite similar for the locally excited states, but again differ considerably for the CT state, a further sign of a severe self-interaction problem in TD-DFT.

Table 3.2: TD-DFT and CC2 vertical excitation energies (in eV, ordered according to the CC2 values), calculated at the DFT optimised CT-state minimum-energy geometry. The excited-state dipole moments (in Debye), as well as the related oscillator strengths (f) in length representation are also given.

		TD-DFT-BP		TD-DFT-B3LYP	
Character		E(eV) (f)	$ \vec{\mu}(\text{ex}) $	E(eV) (f)	$ \vec{\mu}(\text{ex}) $
CT	$\pi \rightarrow \pi^*$	0.671 (0.001)	42.892	1.365 (0.000)	45.715
LE	$\pi \rightarrow \pi^*$	2.729 (0.074)	10.507	3.000 (0.146)	11.865
LE	$\pi \rightarrow \pi^*$	2.493 (0.002)	20.854	2.915 (0.000)	9.192
LE	$n \rightarrow \pi^*$	1.986 (0.001)	7.051	2.817 (0.004)	31.645
PCT	$\pi \rightarrow \pi^*$	2.540 (0.116)	17.379	3.071 (0.079)	12.975

		CC2				
Character		E(eV)(f)	$ \vec{\mu}(\text{ex}) $	μ_x	μ_y	μ_z
LE	$\pi \rightarrow \pi^*$	2.956 (0.000)	46.844	22.87	126.98	-2.26
LE	$\pi \rightarrow \pi^*$	3.114 (0.190)	11.170	8.35	7.40	-0.44
CT	$\pi \rightarrow \pi^*$	3.421 (0.001)	8.863	2.13	8.50	-1.34
PCT	$\pi \rightarrow \pi^*$	3.487 (0.001)	6.043	2.37	5.43	-1.20
LE	$n \rightarrow \pi^*$	3.518 (0.102)	10.766	7.18	8.02	0.29

Since the CT state is energetically well separated from all other states at the level of TD-DFT it was readily possible to determine its minimum-energy geometry by relaxing all degrees of freedom. On going from the ground state to the CT minimum energy geometry significant structural changes do occur, as is shown in Tab. 3.3: the dihedral angles θ_1 and θ_2 (cf. Fig.3.1) within the phenothiazine subunit change by $\Delta\theta_1 = -29.9^\circ$ and $\Delta\theta_2 = -27.4^\circ$, respectively, with the phenothiazine becoming virtually planar in the CT state, while torsional angles between the individual subunits alter by $\Delta\alpha = -14.9^\circ$ (phenyl-phenothiazine) and $\Delta\beta = -8.6^\circ$ (phenyl-isoalloxazine).

Table 3.3: Dihedral angles (torsions) between phenothiazine and phenyl(α), isoalloxazine and phenyl(β), and those responsible for the (non-)planarity of phenothiazine (θ_1 and θ_2). For definition of these angles see Fig. 3.1

torsion	S_0	LE (Δ)	CI (Δ)	CT (Δ)
α	-33.4	-27.6 (5.8)	-38.8(-5.4)	-48.3 (-14.9)
β	95.2	94.1 (-1.1)	90.0 (-5.2)	86.6 (-8.6)
θ_1	28.7	34.0 (5.3)	21.2 (-7.5)	-1.2 (-29.9)
θ_2	26.8	27.5 (0.7)	16.9 (-9.9)	-0.6 (-27.4)

The related changes in the bond distances are compiled in Tab. 3.4. We observe elongations of all C-N and C=O bonds and contractions of all C-C bonds of the isoalloxazine shown in Table. 3.4. The C21-N29 double bond lengthening by nearly 9 pm appears to be by far the most energetic bond (this bond is also important for the Cysteinyl-flavin addition reactions, as shown in the model study by Neiss *et al.*[125]), followed by the two C21-C27 and C21-C22 single bonds, which both become shorter by 4.1pm. The contractions of the C4-S8 and the C3-N31 single bonds in phenothiazine lead to the planarity of this subsystem which is reached at the CT-min structure.

All these bond contractions shown in Tab. 3.4 can already qualitatively be predicted based on the simple orbital picture given in Fig.3.2. It is evident by comparing Tables 3.1 and 3.2 that the energetical order of the LE and CT states (computed at the level of CC2 linear response theory) is interchanged with the CT state now lying *below* the LE state. This is indicative for the presence of a conical intersection (CI) between these two states. TD-DFT, on the other hand, does not show anything alike, drawing an entirely different picture of the photophysics of the system. Furthermore in contrast to CC2, there is no increase in the dipole moment of the CT state at the TD-DFT level as can be seen in Tables 3.1 and 3.2.

Table 3.4: Bond distances which differ from their ground state geometry (Δ) by more than 3 pm at any of the three CT-min, LE-min or CI geometries. For the numbering of the atoms see Fig. 3.1.

bond distance	loc.	S_0	LE-min (Δ)	CI (Δ)	CT-min (Δ)
C21-N29	iso.	1.294	1.362 (6.8)	1.341 (4.7)	1.380 (8.6)
N26-C27	iso.	1.381	1.412 (3.1)	1.406 (2.5)	1.418 (3.7)
C22-N35	iso.	1.298	1.327 (2.9)	1.320 (2.2)	1.332 (3.4)
C36-O37	iso.	1.210	1.235 (2.5)	1.230 (2.0)	1.239 (2.9)
C21-C22	iso.	1.469	1.436 (-3.3)	1.445 (-2.4)	1.428 (-4.1)
C21-C27	iso.	1.505	1.472 (-3.3)	1.481 (-2.4)	1.464 (-4.1)
C 3-N31	pht.	1.399	1.378 (-2.1)	1.383 (-1.6)	1.368 (-3.1)
C 4-S 8	pht.	1.784	1.756 (-2.8)	1.763 (-2.1)	1.742 (-4.2)

3.3.3 At the conical intersection structure

In order to map the low-energy region of the conical intersection seam between the LE and CT states we choose to explore the CT surface, which is well separated from all other states at the TD-DFT level and therefore much easier to treat when it comes to geometry optimisations. Of course, a more natural choice would have been to follow the LE state, however geometry optimisations of that state turned out to be difficult due to mixing with other states. Note that the strongly absorbing LE state at TD-DFT level is not the S_1 or S_2 , but rather the S_8 or S_5 state, depending on the functional in use, which, in addition, are surrounded by weak $\pi\pi^*$ artifact states of TD-DFT, the latter occurring at much higher energy at the CC2 level. As a consequence, TD-DFT geometry optimisations of the LE state are hampered by root flipping, which leads to severe convergence problems.

A 3D-subspace \mathcal{V} within the space of all nuclear degrees of freedom was then defined, spanned by three fixed linear combinations of bond distances (stretch), bond angles (bend), and dihedral angles (torsion), respectively (for technical reasons, mixed linear combinations between bond angles and e.g. dihedral angles could not be used). To specify these three fixed linear combinations the TD-DFT gradient of the CT state surface at the GS minimum energy structure was projected onto the subspaces of the bond distances, the bond angles, and the dihedral angles, respectively. The resulting three projections of the gradient so obtained specify the basis of our subspace \mathcal{V} , the $3N - 9$ remaining coordinates are determined such that they form the orthogonal complement to \mathcal{V} . About hundred geometries interpolating between the ground state equilibrium and the CT minimum-energy structure within \mathcal{V} were then generated by relaxing all coordinates in the orthogonal complement to \mathcal{V} . Note that both the ground state equilibrium structure (by construction) as well as the CT minimum-energy structure lie within \mathcal{V} (since the coordinates of the orthogonal complement are relaxed on the CT surface). For all these points CC2 excitation energies of the CT and LE state then were computed generating two surfaces (depicted in Fig. 3.3(a)), from which the conical intersection seam and its lowest energy point were deduced.

In the following we shall refer to the geometry of the minimum of the CI seam as the "CI" geometry. It is evident from Table 3.3 that the phenothiazine at the CI geometry still is non-planar (the θ -angles are well above zero!). The other two twist angles joining the subunits lie in between their values at the GS and the CT minimum, and the same applies also to the bond lengths compiled in Table 3.4.

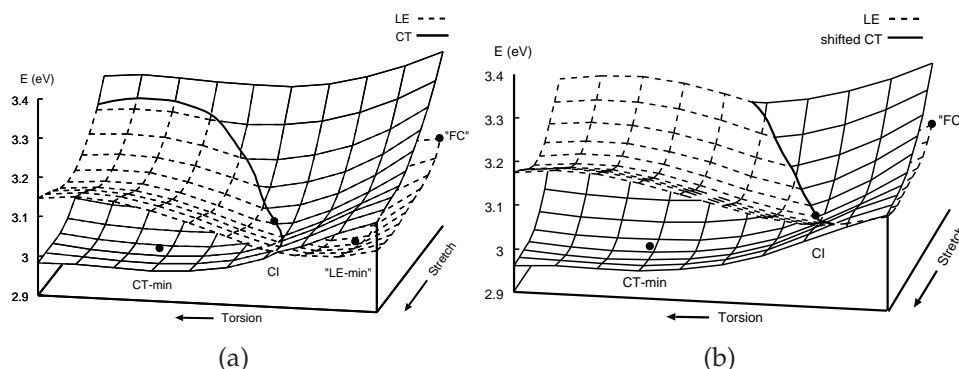


Figure 3.3: LE and CT surfaces within the 2D subspace of \mathcal{V} containing the CT-min and CI structures (bend angle is set to 4°). The ground state minimum geometry has a bend angle of 0° and does not lie within this 2D subspace, hence the *projection* of the Franck-Condon point onto this subspace, indicated by "FC", is given instead (a). The effect of the relative stabilisation of the CT state in the acetonitrile solvent cluster (by 0.068 eV, *vide infra*) on the location of the CI seam is demonstrated in (b) by simply shifting the CT surface relative to the LE surface to lower energies by this amount (cf. Table 3.5).

3.3.4 At the "LE-minimum energy" geometry

The "LE-minimum energy" geometry does not lie within \mathcal{V} , since the nuclear degrees of freedom in the orthogonal complement to \mathcal{V} are relaxed for the CT and not the LE state. Hence the actual LE-minimum lies energetically a bit below its projection onto the \mathcal{V} subspace. Nevertheless, we denote the projection of the LE-minimum onto \mathcal{V} as the "LE-minimum" for the following discussion.

Considering Fig. 3.3(a) one can conclude that (i) there exists a conical intersection seam in the low-energy region of the CT and LE surfaces, (ii) the downhill path from the Franck-Condon (FC) point on the LE surface to the LE-minimum does not lead directly through the conical intersection seam, (iii) the LE-minimum and the CT-minimum are separated by a comparatively low barrier of less than 1 kcal/mol, perhaps somewhat larger taking into account that we only consider the LE-minimum projected onto the \mathcal{V} subspace.

Hence, the question to what extent the CI seam is responsible for fast transfer of the population from the bright LE state to the dark CT state can not conclusively be answered from this picture. However, one has to realize that this is the situation in the gas phase, while all experimental measurements so far have been performed in solution. It is anticipated that solvent effects dramatically stabilise the CT state (due to its large dipole

moment) relative to the LE state, primarily due to dipole-dipole interactions for polar molecules, but also due to inductive interactions in less polar solvent environments. As a result, the conical intersection seam is expected to be shifted in position towards the Frank-Condon point on the LE surface.

Since the two surfaces are quite close already at the FC point a relevant question to be answered is if due to solvent effects the CT surface now moves *below* the LE surface at the GS minimum structure, such that the CI seam now is *above* the FC point, becoming irrelevant for the subsequent photophysical processes. In order to describe solvation effects, computationally rather inexpensive continuum model approaches are frequently employed. However, the results from continuum model calculations may sensitively depend on the specification of the molecular cavities, i.e., the radii of the spheres used for their construction. In particular, spheres used for atoms in ions or zwitterions should be specified differently from those of atoms in neutral molecules [176].

In order to avoid any arbitrariness caused by different specifications of such spheres for the LE and CT states we performed instead calculations for a solvent cluster including the dyad solute surrounded by 20 acetonitrile solvent molecules. Such an approach is of course much more expensive and became only feasible recently by virtue of recent advances in local correlation methods [106].

3.3.5 Solvent Effects

Local CC2 excitation energies for the LE and the CT states were calculated for the dyad in the GS minimum geometry within a cluster of 20 acetonitrile solvent molecules. The geometry of the solvent cluster was generated according to the following three steps: (i) The dyad plus eight nearest acetonitrile molecules were optimised at the DFT level, in order to relax the dyad in the presence of a few solvated molecules. (ii) The dyad structure so obtained was solvated with 480 acetonitrile molecules, the structure and positions of these solvent molecules then optimised by using the classical force field implemented in the MOLOC program [177, 178]. The structure of the dyad solute molecule was kept frozen in this process. (iii) All solvent molecules within a radius of 9 Å from the midpoints of each chromophore (i.e. the midpoints of the middle rings in isoalloxazine and phenothiazine, respectively) were selected and included in the subsequent local CC2 calculations.

The resulting solvent cluster shown in Fig. 3.4 comprises 20 acetonitrile

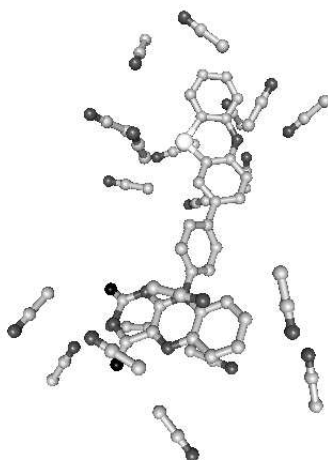


Figure 3.4: Solvent cluster consisting of the dyad (solute) and 20 acetonitrile (solvent) molecules

solvent molecules. This cluster of course reflects just one particular low-energy configuration of the heterogeneous solvent environment. However, in order to answer our rather qualitative question about the order of the two states in the presence of a solvent environment we think that this is a viable approach. The corresponding local CC2 calculations are still rather expensive and belong, to our knowledge, to the largest excited state calculations performed so far at a level beyond TD-DFT. The TD-DFT and LCC2 results obtained for this solvent cluster are compiled in Table 3.5.

It is noteworthy that the CIS method is not able to find an excited state with dominant charge transfer character. Our calculations show that none of the states with a reasonable HOMO-LUMO singles contribution (essential for considering the corresponding excited state as the CT state) in their excitation vector exhibit a sufficiently large excited state dipole moment. This holds true for the isolated dyad, as well as for the dyad solvent cluster. Nevertheless, the domains provided by the CIS wavefunction for the subsequent LCC2 treatment are sufficient to pick up and properly describe the CT state (this was already found to be the case in Ref. [106]).

As can be seen from Table 3.5 the solvent environment shifts the excitation energy of the LE state from 3.160 eV to 2.798 eV. A further shift of 0.13 eV to the red is observed on going from the cc-pVDZ to the bigger aug-cc-pVDZ basis (isolated dyad). Taking this into account we end up at an excitation energy of 2.67 eV for the $\text{LE} \leftarrow \text{S}_0$ excitation in the solvent cluster, which is not too far from the red onset of the LE absorption band at 2.53 eV, measured by Schneider *et al.*[30]. The omission of the side chain and

methyl group has only a minor effect on the excitation energies, as already shown before [106]. The excitation energy for the CT state, on the other hand, is shifted from 3.550 eV to 3.120 eV, leading to a net stabilisation of the CT state relative to the LE state of 0.068 eV, which is far from the value required to switch the order of the states at this geometry (note that we have taken the most polar solvent used in the experiment). Nevertheless, already a constant shift of the CT surface relative to the LE surface by -0.068 eV is sufficient to remove the original barrier between the LE- and CT-minima entirely such that the CI seam now comes into play, as is shown in Fig. 3.3(b).

One can further anticipate that close to the CT minimum, where the solvent molecules are allowed to re-orient according to the bigger dipole of the CT state, the stabilisation of the CT- relative to the LE state is much higher than this constant offset of 0.068 eV. Therefore, the CI-seam moves much closer to the FC point than Fig. 3.3(b) is suggesting. We can conclude from our LCC2 calculations on the solvent cluster, that (i) the CI seam remains in the low-energy area of the LE and CT surfaces, particularly below the FC point and hence plays an important role in transferring population from the bright LE to the dark CT state, and (ii) that the solvent environment plays an important role for the efficiency of the population transfer via the conical intersection. It is of course anticipated that close to the CT minimum, where the solvent molecules are allowed to re-orient according to the bigger dipole of the CT state, the situation is entirely different, substantially stabilising the CT state relative to the LE state, as already stated above. We can conclude from our LCC2 calculations on the solvent cluster, that (i) the CI seam remains in the low-energy area of the LE and CT surfaces, particularly below the FC point and hence plays an important role in transferring population from the bright LE to the dark CT state, and (ii) that the solvent environment plays an important role for the efficiency of the population transfer via the conical intersection.

It is worth mentioning that the TD-DFT calculations, on the other hand, predict a relative stabilisation of the LE state relative to the CT state by 0.141 eV (even though the dipole moment of the CT state is even larger than that predicted by CC2), thus quite the opposite to the LCC2 calculations and also quite contrary to what one would expect from physics. This again illustrates the problems TD-DFT has in describing CT situations.

Table 3.5: Excitation energies (eV), oscillator strengths (length representation) and excited state dipole moments (debye) for the dyad and the dyad solvent cluster (dyad plus 20 acetonitrile molecules). For the local CC2 calculation only the excitation energies are given. ΔLE and ΔCT are the solvatochromic shifts (eV) obtained as the differences of the excitation energies in the gas phase and the solvent cluster environment.

structure	state	TDDFT-B3LYP			LCC2
		E(eV)	f	$ \vec{\mu}(\text{ex}) $	E(eV)
dyad	LE	3.029	0.189	13.018	3.160
dyad+20 ACN	LE	2.807	0.1715	28.11	2.798
dyad	CT	1.834	0.002	46.034	3.550
dyad+20 ACN	CT	1.753	0.0068	54.52	3.120
ΔLE		-0.222			-0.362
ΔCT		-0.081			-0.430
$\Delta\text{CT}-\Delta\text{LE}$		0.141			-0.068

3.4 Conclusions

In this work the photophysics of the phenothiazine-phenyl-isoalloxazine dyad synthesised and characterised recently by Daub and coworkers were investigated by using a combination of TD-DFT and Coupled Cluster Response Theory (CC2 model). We have located a conical intersection seam in the low-energy region of the potential energy surfaces of a bright locally excited (LE) and a dark charge-transfer (CT) state. The question, if this conical intersection seam leads to fast transfer of the population from the excited LE to the dark CT state remains inconclusive for the molecule in the gas phase. However, for the solvated dyad, the CT state is expected to be strongly stabilised by the solvent environment relative to the LE state with much smaller dipole moment. In order to show that the solvent environment is not yet shifting the conical intersection seam *outside* of the low-energy region, i.e., above the FC point, we performed local CC2 calculations on a solvent cluster involving the dyad plus 20 additional acetonitrile molecules.

Based on these calculations we propose that for the solvated dyad fast population transfer indeed occurs from the bright LE to the dark CT state via the conical intersection seam, which hence is responsible for the observed fast fluorescence quenching and low quantum yield observed in the experiments.

Chapter 4

BLUF DOMAINS

4.1 Introduction

The first BLUF domain, which is the subject of this study, was discovered in a flavoprotein AppA, a transcriptional anti-repressor from the purple photosynthetic bacterium *Rhodobacter sphaeroides* and was shown to control photosynthesis gene expression depending on blue-light and/or oxygen conditions[39]. Since then several BLUF signaling domains have been found in prokaryotic and eukaryotic microorganisms all having common spectroscopic features for their signaling state (see below), although having different physiological functions such as photophobic response in photo-activated adenylyl cyclase (PAC) of *Euglena gracilis*[40].

Apart from large conformational changes which lead to the biological signal, an interesting and so far not fully understood aspect of such a receptor domain is the extent and nature of the slight and yet profound conformational changes in the FAD chromophore and its neighborhood upon blue-light illumination. In the case of BLUF domains the absorp-

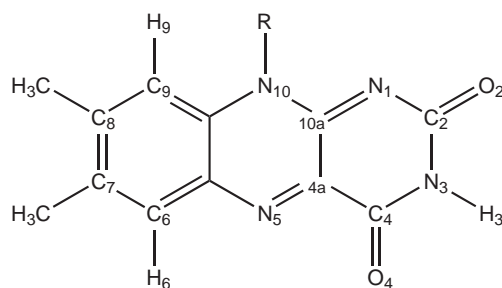


Figure 4.1: Isoalloxazine ring of FMN

tion spectra of the photoinduced (signaling) state is only 10nm red shifted compared to the dark adapted (receptor) state[39] stressing the fact that the chromophore itself does not change its conformation as such. This indeed is the common feature of all BLUF domains known to this date as is evident from the extensive spectroscopy measurements carried out on AppA[179, 180, 181, 182, 183], BlrB[34, 184], Slr1694[185, 186, 187, 188] and Tll0078[189, 190, 191] so far. The signaling state which remains stable for minutes[39], is formed within 1ns[47, 192] after photo-excitation without an intermediate which is stable at room temperature. The hydrogen bonding network associated with this orientation proves to be crucial for the signaling state formation. FTIR and Raman Spectroscopy measurements carried out on the dark- and light-adapted states AppA[193], Slr1694[187] and Tll0078[191] show a 10cm^{-1} red-shift in the stretch mode assigned to the $\text{C4}=\text{O4}$ carbonyl group of Flavin-Mono-Nucleotide (FMN, see Fig.4.1). This result is interpreted as the formation of a stronger hydrogen bond at O4 position of the flavin chromophore in the signaling state.

A conclusive mechanism for the signaling state formation in BLUF domains is complicated by the fact that there is no unique crystal structure available for neither the dark, nor the signaling state. There is some ambiguity regarding the position and/or orientation of the Gln-51, Trp-92 and Met-94 with respect to the flavin moiety. (cf Fig.4.2, the numbering throughout this paper is taken from the crystal structure of BlrB[34]).

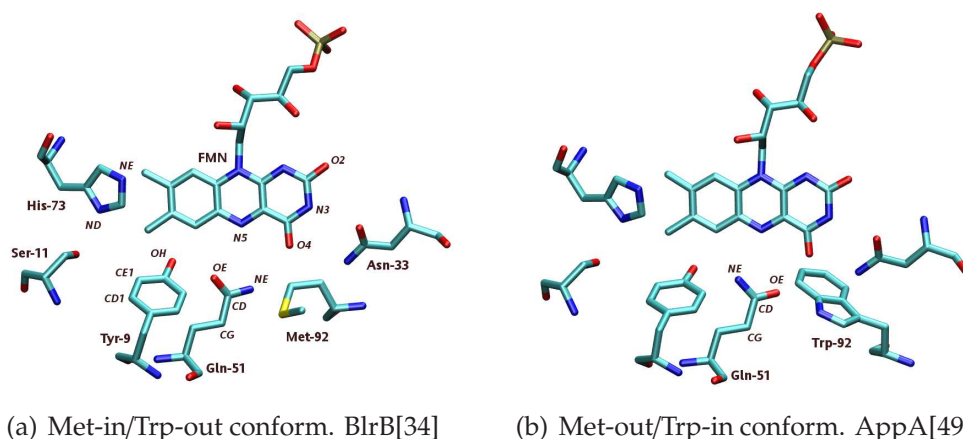


Figure 4.2: Environment of Flavin cofactor (FMN) in the different Met-in/Trp-out and Met-out/Trp-in conformations as found in the x-ray structures of BLUF domains. The terms Met-in and Trp-in are used in the following for the conformations shown in Fig. 4.2(a) and Fig. 4.2(b) respectively.

An electron transfer process from Tyr-9 to FMN has been proposed to be responsible for the fluorescence quenching of the flavin moiety[194, 180, 195, 196, 191, 197, 188]. A mechanism postulated by Gauden *et. al.* involves a light-driven electron and proton transfer from Tyr to FAD followed by a hydrogen bonding network rearrangement in AppA[183] and Slr1694 mainly due to the rotation of Gln[47]. Recently Domratcheva and coworkers postulated the possibility of Gln rotation during the lifetime of biradical state of the system[48] which is then followed by biradical recombination leading to the tautomeric form of Gln-51. The possibility of tautomerization was proposed earlier by the FT-IR study of Stelling *et. al.* on AppA[46]. The latter group has detected a transient infra-red absorption band at 1666cm^{-1} , only present in the signaling state, which was assigned to the carbonyl stretch of Gln-51.

In this study we present a plausible mechanism for the formation of the signaling state after the initial photo-excitation based on extensive Quantum mechanics/Molecular Mechanics calculations (for a recent review see Ref.[59]) with appropriate QM methods describing ground- and electronically excited states. In these calculations the FMN chromophore, along with its nearest neighboring amino acids, were treated quantum mechanically (QM part), while the remaining protein environment was treated at the level of a classical force field (MM part). The effect of the MM environment enters the QM Hamiltonian at the level of distributed point charges. There is strong evidence, that after photo-excitation of a Locally Excited (LE) state the Tyr9-FMN Charge Transfer (CT) state is populated via a conical intersection. Our calculations show that the downhill path on this CT surface then leads via a further conical intersection with the closed shell ground state to the biradical species (Tyr \bullet -Flavin \bullet), hence confirming the experimental results from the theoretical perspective.

Furthermore, possible channels for the recombination of the biradical are explored. The barrier for the process involving rotation of the Gln-51 in the presence of Tyr \bullet and Flavin \bullet is compared to the barrier for rotation-free direct radical recombination to Gln-Oxygen (see Fig. 4.8). The latter possibility has not been considered so far, yet it is demonstrated in this work that it is more compatible with the short lifetimes observed in spectroscopy measurements than alternative recombination mechanisms.

Finally, the controversial role of Met-94 and Trp-92 residues (see Fig. 4.2) for the signal transduction is investigated. Assuming that the signaling state (to which the observed 10nm red-shift in absorption spectrum is assigned) is generated by tautomerization of the Gln moiety[46, 48], which is also one of the conclusions of the present work, the influence of protein conformations, differing in the position of Met-94 and Trp-92 residues, on

the relative stability of Gln-tautomer are considered. Calculations were carried out on the two Met-in (or Trp-out) and Met-out (or Trp-in) conformations, for which the terminology *Met-in* and *Trp-in* (see Fig. 4.2(a) and 4.2(b)) will be used throughout this paper. Based on the results obtained we conjecture that the Trp-in form corresponds to the final signaling state of BLUF domains.

4.2 Computational Details

The monomer A of the crystal structure of BlrB[34] (PDB code: 1BYC) was used to setup the Quantum Mechanics/Molecular Mechanics (QM/MM) calculation. A pre-equilibrated water droplet (radius 20Å) containing 5187 water molecules was used to solvate this initial structure. The droplet boundary of 2.5Å was constrained with a quartic force of 0.2 [24 kcal/mol/Å²] using the Miscellaneous Mean-Field Potential (MMFP). All bonds to hydrogen atoms were held by SHAKE[56]. For FAD we used the parameters available from CHARMM27 nucleic and amino acids[66, 67, 68, 69, 70](see Supporting Information). The system was heated from 50K to 300K using a 1fs time step and 10 K temperature increase for every 25th time step during which the backbone was constrained in a harmonic potential. This constraint was then lifted gradually and the system was thereafter allowed to relax at 300K over 500000 time steps (\approx 500ps). Simulations showed a large mobility of Tyr-9 next to Gln-51 and FAD (see Results for further details).

We found a buried pocket between Tyr-9 and Ser-11, which may be transiently occupied by a water molecule (denoted as WAT, see Fig. 4.3) not observed in the X-ray structure analysis. Simulations with this additional buried water molecule stabilized the conformation of Tyr-9 in the same orientation as observed in the reference structure 1BYC. The input geometries for the QM/MM calculations were prepared by selecting representative structures from the Molecular Dynamics (MD) run and minimizing the selected structure using the force field.

The QM region naturally comprises FMN as the main chromophore and Tyr-9 since a Charge Transfer (CT) process from Tyr-9 to FMN appears to play an essential role in the mechanism of the photo-cycle. Additionally, Gln-51 as a further key player, His-73 (protonated at NE, see Fig.4.2(a)), the WAT molecule and its neighboring Ser-11 were also included in the QM region (see Fig. 4.2(a)). The His-73 moiety, being rather close to FMN, might also be directly involved in the relevant excited states, or at least significantly interact with FMN and Tyr-9 and therefore influence the energetic

position of the CT state. Additionally, for calculating harmonic vibrational frequency and vertical excitation energy calculations at the proposed dark, intermediate and signaling state structures, the above mentioned QM region was augmented by Asn-33 and Trp-92. In order to explore the possible different pathways for biradical recombination a somewhat reduced QM region was employed, consisting of only FMN, Tyr-9, and Gln-51.

To study the possibility of Gln-51 tautomerization (imide form of Gln-51) and the orientation exchange of the Trp-92 and the Met-94 residues (leading to the Trp-in and Met-in conformations respectively) additional QM/MM calculation based on protein model systems were carried out. To this end the wild-type BLUF-FAD complex and the imide were created on the basis of the X-ray structure of 1BYC. Due to the lack of any reliable Charmm parameters for the imide tautomer of Gln-51 these models were initially minimized by using the force field MAB as implemented in the MOLOC program[177, 178]. Thereafter, MD simulations of 10ps were performed. An H-bond weight of 1.78 was applied, which corresponds to the H-bond strength parameterized for describing the intramolecular H-bond enthalpy rather than for the intermolecular ligand binding H-bond free energy (default in MAB).

Furthermore, a second set of structures with a different orientation of the loop before $\beta 5$ was generated. This loop includes the Trp 92 and Met 94 units. In the X-ray structure 1BYC the residue Met94 is oriented towards Gln-51 next to FAD and Asn-33 while Trp-92 is exposed to the solvent (see Fig.4.2(a)). The second conformation with Trp-94 pointing in and Met-92 pointing outwards towards the solvent was created by adopting the relevant environment of the X-ray structure 1YRX of the BLUF-domain of AppA and 2HFN (chain D) of Slr1694 (see Fig.4.2(b)).

The QM/MM interface ChemShell[72] was used throughout in this study. The QM/MM coupling was calculated using the charge-shift scheme and link atoms[72]. For the MM part of the QM/MM calculations the DL_POLY molecular dynamics package[71] (with CHARMM force field parameters) was applied. The HDLCopt optimizer[198] implemented in ChemShell was employed for the QM/MM geometry optimizations, which has to be supplied with the incremental energy and the incremental nuclear energy gradient of the QM part.

Geometry optimizations on the potential energy surface of the electronic ground state were carried out at the level of Density Functional Theory (DFT) employing the B3LYP [158] hybrid functional, and, for purpose of verification, also at the level of Local-Møller-Plesset perturbation theory of second order (LMP2)[199]. For these energy functionals efficient analytic energy gradients are available.

Geometry optimizations on the potential energy surfaces of the electronically excited states, i.e., the charge-transfer surface, (due to the lack of alternatives for extended systems of this size), were performed at the level of Time-Dependent DFT (TD-DFT) response theory, applying the analytical TD-DFT energy gradient program of Furche *et. al.*[109]. For that purpose the ChemShell interface had to be modified accordingly. It is well known that TD-DFT grossly underestimates excitation energies of CT states (cf. Ref. [114] and references therein). This problem is directly related to the electronic self-interaction inherent in DFT[164, 165] and can partly be alleviated by employing hybrid functionals with a larger fraction of Hartree-Fock exchange. The B3LYP functional (and naturally even more so the BP functional) turned out to dramatically underestimate the excitation energies of all kinds of CT states of the system (cf. Table 4.1), which in turn would mix with the locally excited state of interest. Therefore, instead of B3LYP the BHLYP functional [200] with a larger fraction of exact Hartree-Fock exchange was employed for the TD-DFT excited state geometry optimizations.

For individual TD-DFT optimized geometries the related excitation energies were also computed at the level of Time-dependent Coupled Cluster response theory, utilizing the CC2 model [105, 168]. This strategy of combining TD-DFT optimized geometries with CC2 response single point excitation energy calculations was successfully employed by us recently in a study on the photophysics of a phenothiazine-phenyl-isoalloxazine dyad, where an electron transfer process from phenothiazine to isoalloxazine is activated upon locally exciting the flavin chromophore [114]. The electronic ground state calculations on the neutral biradical along the different possible pathways of biradical recombination were carried out at the level of spin-unrestricted DFT. Here, in order to save computational resources, the B3LYP functional and the SVP basis set were used, which should do for that purpose.

For the (TD)-DFT and CC2 calculations, we used the TURBOMOLE[154] program package using the def-TZVP[201] and cc-pVDZ basis sets respectively. The Local-MP2 calculations were performed with the MOLPRO[173] program employing the aug-cc-pVDZ[202] basis set.

4.3 Results and Discussion

4.3.1 Search for a representative structure

The first X-ray crystal structure of the dimer of the N-terminal AppA 17-133 BLUF-domain was determined by Anderson *et. al.*[49] and exhibits an $\alpha + \beta$ sandwich with the FAD aromatic moiety bound in between the two long helices of a ferredoxin like fold. The dimer interface however is distorted by detergent molecules.

The NMR solution structure of AppA BLUF domain was determined by Grinstead *et. al.*[203] and shows essentially the same dimer structure as in the X-ray crystal structure, despite the fact, that the orientation of the key Gln-51 residue and a C-terminal loop containing Trp-92 near flavin is not resolved. The structure of the light induced signaling state of AppA BLUF domain is currently not known, but upon irradiation some residues show changes in the NMR-chemical shifts, most of them located near or in the C-terminal region[204]. This dynamic behavior found in the NMR solution structure suggests that the C-terminal part of the BLUF domain is involved in the signaling and may sample different conformations.

Jung *et al.* also reported a crystal structure of mutated AppA 1-124 C20S BLUF domain in the dark state[35]. This structure shows a different orientation of the C-terminal stretch with a Met residue oriented towards Gln (see Fig. 4.2(a)) and the Trp-residue pointing out to the solvent. Similar orientations for the Trp and Met residues are observed in the crystal structures of BlrB[34], Tll0078[189] and Slr1694[185]¹. In the following we will denote such conformations as *Met-in*, whereas for the crystal structure proposed by Anderson *et. al.*[49] the term *Trp-in* will be used (see Fig. 4.2(b)).

The orientation of the Gln-51 moiety in the dark state has also been the subject of some debate. In the *Met-in* crystal structures the amide group of Gln-51 acts as a hydrogen bond donor to the FMN-N5 whereas the C=O group is the acceptor of a hydrogen bond from Tyr-9-OH (see Fig. 4.2(a)). In the *Trp-in* structures, on the other hand, the carbonyl group of Gln-51 is rotated about the CG-CD bond by 180° as shown in Fig. 4.2(b). As a structural basis for our theoretical study we adopted the high resolution dark state structure of the well characterized BlrB BLUF protein solved by Jung *et. al.*[34].

Of primary interest in the present context are the photo-induced processes involved in the formation of the signaling state, taking place in the active

¹Only in one monomer (D) of Slr1694 we find the Trp-92 inside the protein and next to Gln-51, however there seems to be no H-bond interaction between these two residues

site of the BLUF domains in the vicinity of FMN cofactor. Since this active part of the protein shows a strong conformational similarity for all structurally known BLUF domains and furthermore the structural rearrangements are small on going from the dark/receptor to the light/signaling state (as apparent from the small 10nm shift in absorption spectrum) it appears to be sufficient to consider one representative structure.

In our initial MD simulations a significant instability in the orientation of the Tyr-9 moiety relative to the flavin cofactor was observed (cf. Fig. 4.3). The Tyr-9 residue appears to switch between two distinct orientations (colored red and magenta in Fig. 4.3). This switching motion is modulating the interaction of the environment with the FMN chromophore.

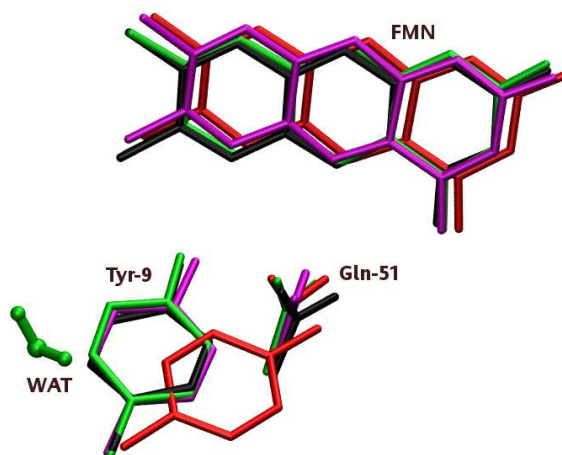


Figure 4.3: The crystal structure of BlrB, shown in black, is superimposed on three configurations obtained from molecule dynamics simulations. The conformation shown in green corresponds to the case where an additional water molecule is placed within the cavity found near the Y9 residue. In the conformation shown in red, the Tyr-9 residue is severely dislocated relative to the crystal structure, while the conformation displayed in magenta coincides closely with the experimentally resolved structure. The MD simulations predict the magenta conformation as energetically more favorable than the red form.

The energetic position of the CT state is anticipated to be very sensitive to the environment, certainly much more so than the LE state carrying the oscillator strength. Therefore the relative energetic position of LE and CT

state, a key property for the conjectured mechanism, is expected to differ appreciably for the two orientations of Tyr-9. However, only one of the two orientations is in good agreement with all published crystal structures to this date for the BLUF domains, in the other conformation the TYR ring is severely dislocated relative to the crystal structure.

As already pointed out in the previous section an additional water molecule (WAT) was placed into the cavity of the crystal structure between the Tyr-9, His-73 and Ser-11. This was motivated by the fact that the isotropic displacement (B-factor) of the phenyl ring of Tyr-9 is asymmetric and the atoms CD1 and CE1 (shown in Fig. 4.2(a)) oriented towards Ser-11 show a higher mobility. The MD run performed for the BlrB BLUF domain in the presence of WAT, which is primarily interacting with the Tyr-9, exhibits a much more stable orientation of this amino acid leading to conformations in close agreement with the known crystal structures.

Two different types of H-bond interactions between WAT and Tyr-9 were observed, where in both cases WAT acts as the H-bond donor and the Tyr-9 ring or the Tyr-9 oxygen atom as the corresponding acceptor. We employ the notation $Y9-\pi$ and $Y9-OH$ for these two orientations, respectively (see Fig. 4.5(a) and 4.5(b)).

The orientation of the Gln-51 is a further issue: in the crystal structure of BlrB the orientation of Gln-51 is such that the carbonyl group of Gln-51 and the OH group of Tyr-9 form a H-bond. The label $Q51-O\epsilon$ (the $O\epsilon$ and $N\epsilon$ atoms of Gln are shown in Fig. 4.4) is used in the following to denote such an orientation, thereby emphasizing the role of $C=O$ as the H-bond acceptor in this alignment.

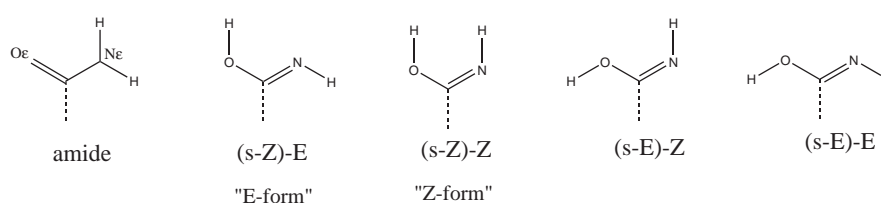


Figure 4.4: Different tautomer/isomers of Gln, nomenclature taken from ref. [205]

The MD simulations do not reveal any tendency for an orientation of Gln-51 different to $Q51-O\epsilon$. This actually also holds true for MD simulations in absence of WAT. This is not unexpected due to the rigidity of the environment around Gln-51 which prevents a rotation of Gln-51 even for the case of a much more fluxional Tyr-9 orientation, i.e., without WAT. Nevertheless, orientations of Gln-51 different to $Q51-O\epsilon$ are frequently discussed in

the literature[49, 204, 203, 47]. Therefore, alternative orientations of Gln-51, generated by systematic rotation of the amide group of this residue were studied in the context of this work. Our attempts to find a minimum structure which would correspond to a 180° flipped Gln-51 residue (as shown in Fig. 4.2(b)) failed. Instead a rather stable arrangement was found where the Gln-carbonyl group is pointing away from the Tyr-9 residue which corresponds to a rotation of roughly 90° about the CG-CD bond of Gln-51. The former still acts as H-bond donor, namely to the Gln-N ϵ (see Figures 4.5(c) and 4.5(d)). For these possible conformations the notation Q51-N ϵ is used in the following.

The four geometries displayed in Fig. 4.5 represent the four possible orientations of WAT and Gln-51 relative to the Tyr-9 residue, respectively. The orientations of all the other amino acids in the vicinity of flavin, which are highly conserved in various BLUF domains, remain very similar to those of the crystal structure during the MD simulation runs. After equilibration the four individual geometries of Fig. 4.5 were optimized by using the force field in order to get good starting geometries for the subsequent QM/MM optimizations.

4.3.2 Ground state QM/MM: Determination of the most stable dark state structure

The QM/MM geometry optimizations were carried out by applying LMP2 or DFT-B3LYP for the QM part in combination with Charmm force field for the MM part (*vide infra*). The results are compiled in Table 4.1. The most stable geometry clearly corresponds to the Y9- π /Q51-O ϵ arrangement. The alternative Y9- π /Q51-N ϵ structure with the Gln-51 carbonyl group being rotated by $\approx 90^\circ$ relative to the experimental crystal structure of BlrB, is about 3.5 kcal/mol (LMP2) or 2.3 kcal/mol (DFT-B3LYP) less stable and most likely not a viable candidate for the dark state structure. Furthermore, a preference for WAT to form an H-bond to the Tyr-9 Ring (Y9- π structures) rather than to the oxygen atom of Tyr-9 (Y9-OH structures) is observed. Such hydrogen bonds also occur in the electronic ground state of 2-Naphthol clustered with two water molecules[206]. Based on these results we postulate that Y9- π /Q51-O ϵ indeed represents the dark state (receptor state) structure from which photo-induced processes are triggered. Consequently, all further calculations presented here use Y9- π /Q51-O ϵ as their starting point.

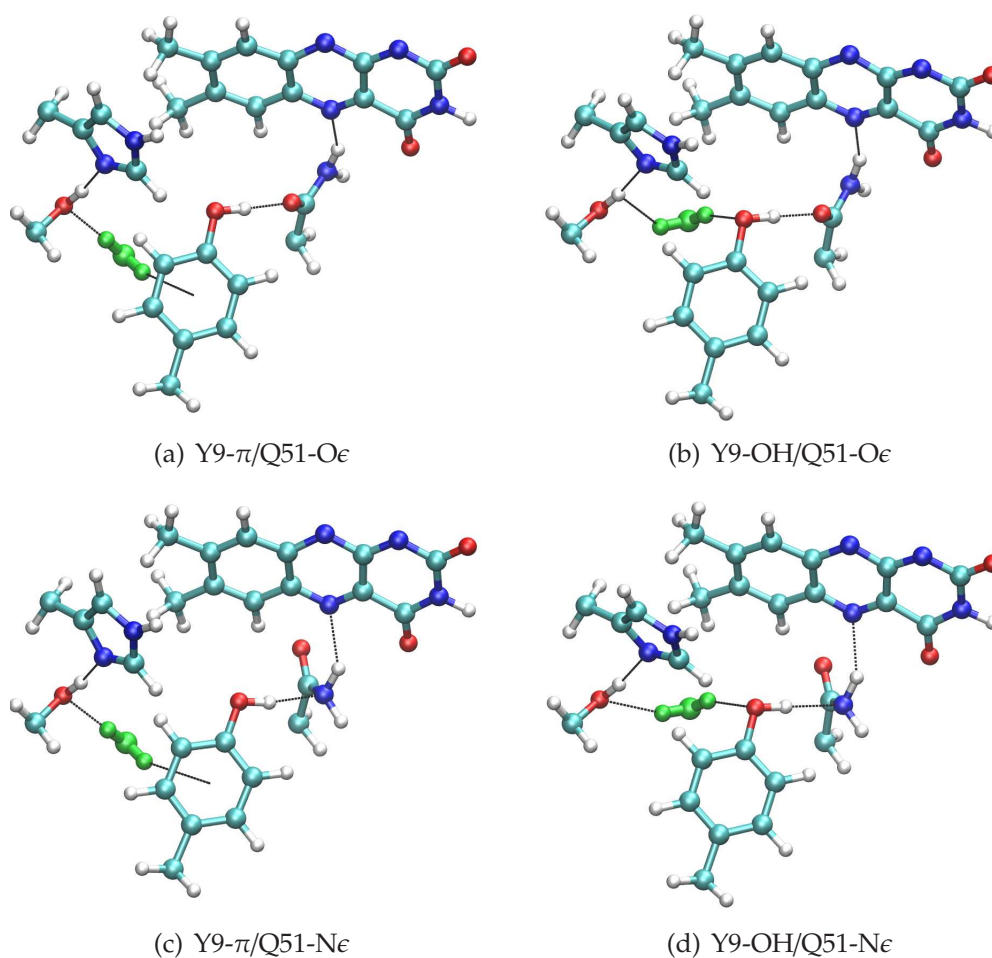


Figure 4.5: Comparison of the ground state structures obtained by MD minimizations. The water molecule (WAT) is a H-bond donor to the ring (Y9- π) or to oxygen atom (Y9-OH) of the Tyr-9 amino acid. The Gln-51 (Q51) is the H-bond acceptor, either through the carbonyl group (Q51-O ϵ) or the amine group (Q51-N ϵ).

Table 4.1: Relative ground state energies (in kcal/mol) of the QM/MM minimum energy structures obtained by LMP2/Charmm and DFT-B3LYP/Charmm, respectively. The QM/MM ground state energy of the Y9- π /Q51-O ϵ structure is taken as reference.

structure	LMP2/Charmm	B3LYP/Charmm
Y9- π /Q51-O ϵ	0.00	0.00
Y9-OH/Q51-O ϵ	3.65	10.84
Y9- π /Q51-N ϵ	3.54	2.25
Y9-OH/Q51-N ϵ	9.13	16.81

4.3.3 Choice of QM method for the excited states

Vertical excitation energies and oscillator strengths for the relevant LE and CT states were calculated with TD-DFT and CC2 response theory at the Y9- π /Q51-O ϵ DFT-B3LYP geometry. The resulting values are compiled in Table 4.2. Related differential densities between ground and excited state densities were also computed with TD-DFT/BHLYP, and plotted in Fig. 4.6.

Table 4.2: Vertical excitation energies (in eV) calculated at the TD-DFT and CC2 levels of theory for the dark structure (Y9- π /Q51-O ϵ). The oscillator strengths, f , are shown in parenthesis.

Character	BP	B3LYP	BHLYP	CC2
LE(FMN)	2.48(0.110)	2.81(0.179)	3.29(0.269)	2.99(0.249)
CT(Y9 \rightarrow FMN)	0.88(0.000)	1.86(0.000)	3.17 (0.001)	3.59(0.000)

For reasons outlined in the section II. we consider TD-DFT results calculated with the BP or B3LYP functionals as entirely un-trustworthy. The largest oscillator strength of 0.269 (TD-DFT/BHLYP) and 0.249 (CC2) was obtained for the LE state of flavin, which must correspond to the experimentally measured absorption bands around 450 nm (2.75 eV).

It is evident from Fig. 4.6 that the differential density of the LE excitation is entirely localized on the isoalloxazine subunit. CC2 response is overestimating the experimental value for the excitation energy by about 0.25 eV, as expected for this method/basis, while TD-DFT/BHLYP is overestimating it by 0.3 eV. The lowest CT state involves transfer of an electron from Tyr-9 to FMN, as is evident from Fig. 4.6. According to the CC2 reference calculations, it lies energetically *above* the LE state for the Y9- π /Q51-O ϵ geometry and carries virtually no oscillator strength.

This is very similar to the situation encountered in a phenothiazine-phenyl-isoalloxazine dyad studied by us previously, where the CT state responsible for fluorescence quenching is populated via the LE state through a conical intersection [114]. For the present system we hence postulate an analogous mechanism to populate the CT state which is corresponding here to a zwitterionic biradical involving the partially positively charged Tyr-9 and the partially negatively charged isoalloxazine subunits, respectively. This

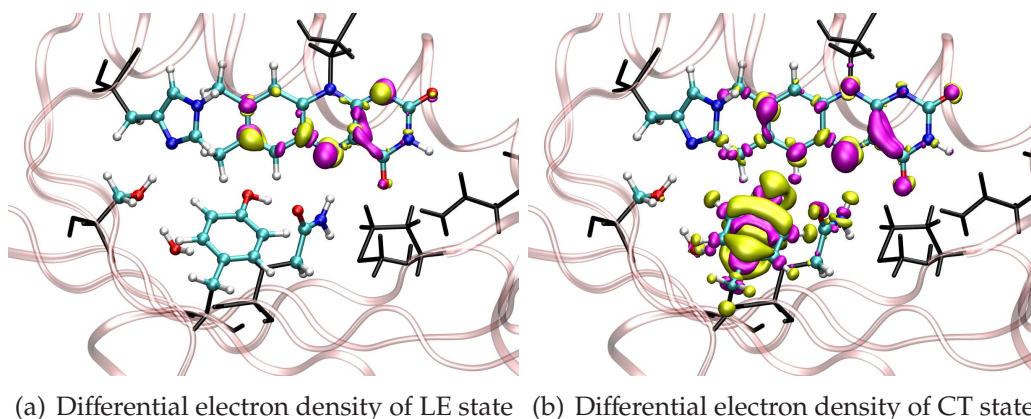


Figure 4.6: Difference in ground- and excited electron densities upon excitation. Yellow and magenta regions represent regions with loss and gain of electron density, respectively. All the graphics presented here has been prepared by the VMD package[207]

hypothesis is supported experimentally by the low fluorescence quantum yield and the short lifetime of the excited FMN measured in BLUF domains[182, 184, 188, 183, 47].

The TD-DFT/BHLYP method does not provide the correct ordering of LE and CT state, and is therefore not able to predict or locate such a conical intersection (in contrast to CC2 response). On the other hand, the TD-DFT/BHLYP excitation energy of the CT state is not nearly as low and wrong as it is for the other two functionals. More importantly, the character and oscillator strength of the LE state provided by TD-DFT/BHLYP closely resembles that provided by CC2 for all relevant geometries, which turns out not to be the case for the other two functionals. Therefore, TD-DFT/BHLYP was used for all excited state geometry optimizations including those on the CT surface.

4.3.4 Charge transfer QM/MM: Biradical formation mechanism

By exploring the CT surface in the vicinity of the Y9- π /Q51-O ϵ ground state minimum geometry it turned out that along the coordinate of proton transfer from Tyr-9 to Gln-51, labeled as P_1 in Fig. 4.7, the total energy of the CT state is decreasing steeply. A sequence of constrained geometry optimizations along this pathway was then performed. The difference of the distance between H and the oxygen atoms of Tyr-9 and Gln-51, respectively, (denoted by Δ_1 , cf. Fig. 4.7) was defined as the constraint. All other degrees of freedom were fully relaxed. On going from the geometry

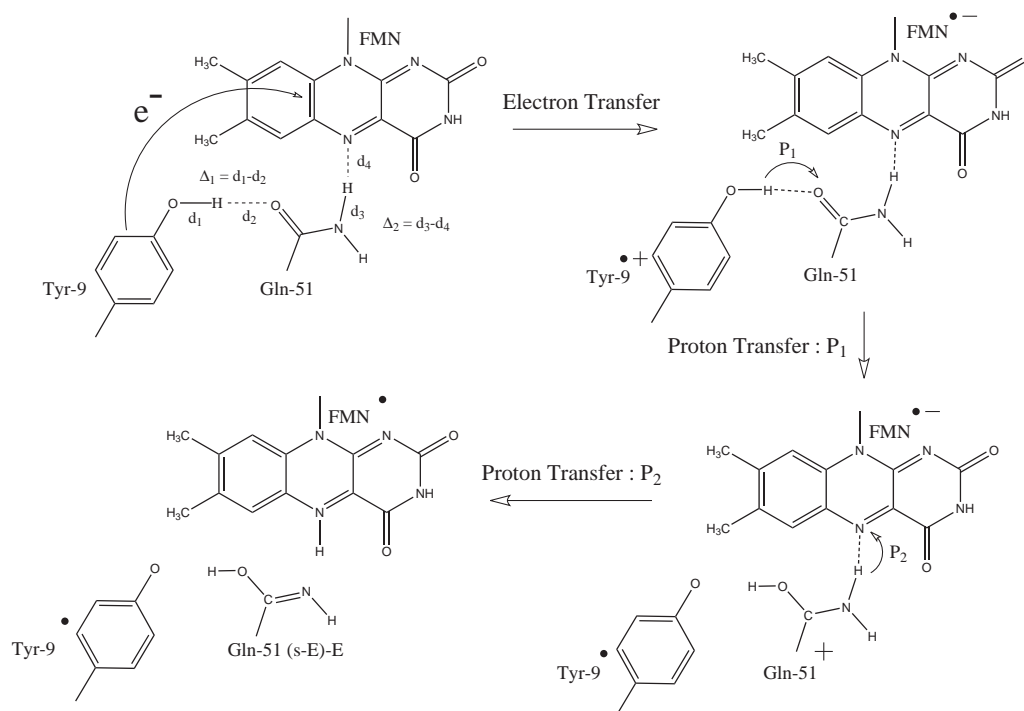


Figure 4.7: Y9•-Q51-FMN• Biradical formation path

corresponding to $\Delta_1 = -0.5$ (proton still on the side of Tyr-9 $\approx \text{Y9}^{\bullet+}\text{-Q51-FMN}^{\bullet-}$ structure) to that corresponding to $\Delta_1 = +1$ (proton transferred to Gln-51 $\approx \text{Y9}^{\bullet}\text{-Q51}^+\text{-FMN}^{\bullet-}$ structure) the CT state drops in energy by 6 kcal/mol (total QM/MM energy of the CT state). The excitation energy alone decreases at the same time from 0.99 to 0.1 eV (TD-DFT, BHLYP) or from 1.41 to 0.31 eV (CC2). This is a clear indication that after the postulated population of the CT state (via LE by conical intersection) the CT state relaxation proceeds via proton transfer (P_1) from Tyr-9 to Gln-51. A full geometry optimization (without constraints) on the CT state surface starting from Y9- π /Q51-O ϵ directly leads to a structure where the proton is fully transferred to Gln-51, implying that this proton transfer proceeds without any barrier in between.

An efficient and accurate way to locate the extremal points on the conical intersection seam requires analytical gradients of the individual excited states and their non-adiabatic coupling vector (cf. see Ref [52] and references therein). For the current system the analytical energy gradients at the CC2 level of theory are clearly too expensive. An alternative strategy used previously for the phenothiazine-phenyl-isoalloxazine dyad[114] involves a combination of TD-DFT geometry optimization and CC2 single

point calculations. Using this approach minimum energy structures on the conical intersection seam can be roughly determined. However this is a rather tedious procedure requiring a very large number of calculations and we have therefore decided not to pursue this issue in the present work.

Nevertheless there is strong evidence that the conical intersection indeed exists: Comparing the Y9- π /Q51-O ϵ and $\Delta_1=-0.5$ structures, the sole appreciable structural change in the flavin binding pocket is the slight displacement of the Tyr-9 proton. One may therefore consider the $\Delta_1=-0.5$ structure to be very close to the Frank-Condon (Y9- π /Q51-O ϵ) structure. Yet the charge transfer excitation energy has dropped dramatically, from 3.59 eV (dark) to 0.99 eV ($\Delta_1=-0.5$) and is energetically already *below* the LE state at the latter geometry. This implies that the Frank-Condon point is located rather close to the conical intersection seam between these two states.

In the process of the first proton transfer (P_1), the distance between FMN-N5 and the closest hydrogen of the Gln-51 amino group is shrinking from 2.15 Å at the Frank-Condon point (Y9- π /Q51-O ϵ) to 1.83 Å at the $\Delta_1 = +1$ geometry. This indicates that the next consecutive step in the relaxation process on the CT surface is a second proton transfer from the Gln-51 amino group to FMN-N5, labeled as P_2 in Fig. 4.7, leading to the Y9 $^{\bullet}$ -Q51-FMN $^{\bullet}$ biradical where the charge separation induced by photo excitation i.e., the zwitterionic character, is equalized (cf. Fig.4.7).

At the Y9 $^{\bullet}$ -Q51 $^+$ -FMN $^{\bullet-}$ geometry the CT excitation energy is already very low, 0.1 eV (TD-DFT, BHLYP) or 0.31 eV (CC2) and it continues to decrease along the second proton transfer coordinate (P_2) in such a manner that a second conical intersection seam between CT and ground state surface is crossed. The closed-shell DFT-BHLYP ground state becomes variationally instable after a certain point along the coordinate is passed and the biradical is taking over the role of the electronic ground state. Time-dependent response methods like TD-DFT or coupled cluster response theory rely on a closed-shell reference wavefunction for the ground state, hence any conical intersection between the electronic ground state and an excited state cannot be dealt with (instead, multi-reference methods with appropriate active spaces would have to be utilized, which however are very cumbersome to use in such extended systems). Hence, no attempt was made to exactly locate the point where the proton transfer pathway dissects the conical intersection seam.

The final Y9 $^{\bullet}$ -Q51-FMN $^{\bullet}$ biradical form was generated by placing the second proton (of Gln-51) at distance of ≈ 1 Å and fully relaxing this structure thereafter by employing spin-unrestricted DFT/B3LYP. Of course the resulting wavefunction describing the biradical is no longer a pure singlet

state, at the Y9[•]-Q51-FMN[•] geometry an eigenvalue of S^2 of 1.05 was obtained, as anticipated for a biradical described by a spin-unrestricted single determinant.

These results all indicate that the CT state, once activated via conical intersection with the LE state, opens a downhill channel within which the first and second proton transfer process (from Tyr-9 to FMN, via Gln-51) occur without any barrier and finally lead to the formation of the neutral biradical Y9[•]-Q51-FMN[•] species. As is evident from Fig. 4.7 the process of generating the neutral biradical Y9[•]-Q51-FMN[•] species involves the tautomerization of the Gln-51 to the imidic form (amide \rightarrow s-E-E, nomenclature taken from ref. [205]. See Fig.4.4) without rotation.

4.3.5 Open-shell QM/MM: Biradical recombination path

Biradical recombination via Gln-51 rotation:

According to the experiments, the Y9[•]-Q51-FMN[•] biradical form has a lifetime of 65ps[47] after which the biradical recombination step is completed. Any proposed process which leads to the loss of the biradical signal must therefore be compatible with a lifetime of this order of magnitude. In order

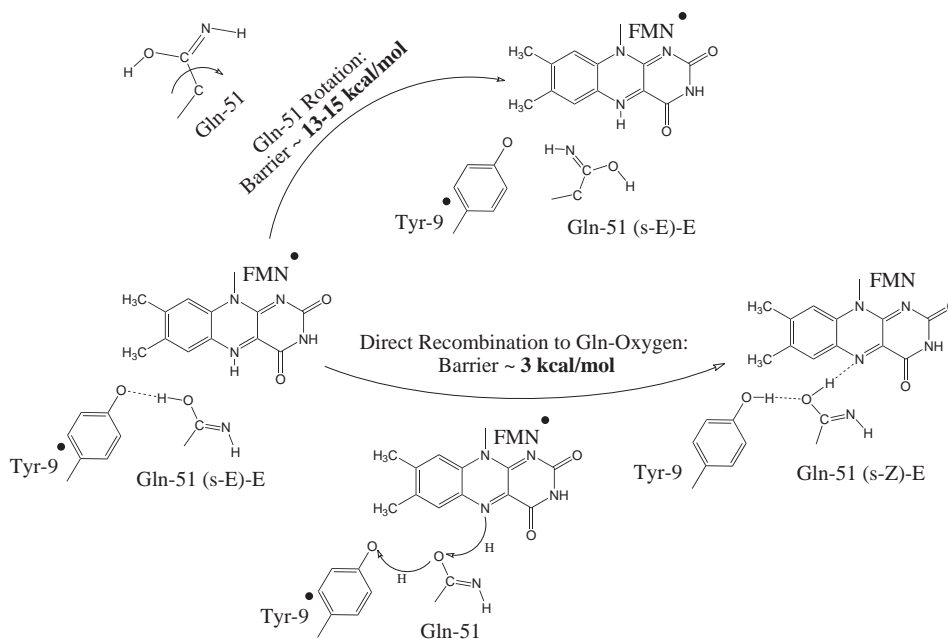


Figure 4.8: Our proposed radical recombination to the oxygen of Gln-51 vs. rotation of Gln-51 in the biradical Y9[•]-Q51-FMN[•] state prior to radical recombination.

to examine the possibility of Gln-rotation as suggested by Domratcheva *et. al.*[48], we carried out a series of constrained geometry optimizations of the neutral biradical state along the rotation angle about the C-C bond of Gln-51 (see Fig. 4.8). These calculations were again performed by applying spin unrestricted DFT with the B3LYP functional.

Our results indicate that, depending on the direction of rotation, a barrier of 13-15 Kcal/mol has to be surmounted. This high barrier implies a lifetime of the Y9[•]-Q51-FMN[•] biradical form of several ms instead of the 65ps imposed by the experiment. In our view, two major factors contribute to this barrier, which is by far too high to be overcome within nanoseconds upon excitation:

- (i) There is a strong H-bond, namely Tyr9-O[•]/Gln51-OH and a weak interaction between Met-94-S and Gln-51-NH₂ which needs to be cleaved during the process of rotation.
- (ii) There are four amino acids around the Gln51 residue, namely Leu-42, Ile-53, Ile-67 and Met-94, which are highly conserved and block rotation of Gln-51 via steric hindrance. For the complete conservation pattern of the BLUF family see the supporting information. Moreover, there are additional rotations required (about the NH double and OH single bonds of Gln-51) in order to get the particular geometry, which, according to Domratcheva *et. al.*, is finally enabling biradical recombination. Appreciating all this, it appears as unlikely that the radical recombination proceeds via rotation of Gln-51.

Biradical recombination without Gln-51 rotation:

An alternative to the rotation of Gln-51 is the formation of the imidic form of Gln-51 (s-Z-E form). We postulate that the hydrogen atom on FMN-N5 is transferred to the Gln-51 oxygen directly. Our calculations predict a barrier of about 3 Kcal/mol for such a process (Fig. 4.8) which involves the isomerization (s-E-E → s-Z-E, these isomers are shown in Fig.4.4) of Gln-51. This barrier indeed matches nicely with the experimental lifetime of 65ps of the biradical state (*vide supra*). Moreover, a much smaller distortion in the conformation of the protein as a whole, namely a small change in the alignment of the hydrogen donor/acceptor groups (i.e. from FMN-N5/Q51-N ϵ to FMN-N5/Q51-O ϵ) is sufficient to allow this step for taking place. Comparing the aforementioned barriers we conjecture that the rotation of the Gln-51 moiety is not a competitive step and therefore unlikely to be of relevance for the mechanism of signaling state formation.

4.3.6 Path to the Signaling State

As pointed out in the introduction the redshift of the IR absorption band assigned to the stretch mode of the C4=O4 bond is anticipated to be caused by the formation of a new hydrogen bond at the O4 position of FMN. Such a hydrogen bond was postulated to be formed after the rotation of the Gl-51 residue [193, 208, 48, 47, 209, 39]. So far we have shown that the structure with Gln-51 in its s-Z-E form, denoted as Q51-E, which is immediately reached upon recombination does not involve any rotation of Gln-51. In the geometry Q51-E, however, no new H-bond to FMN-O4 can be formed. Therefore, alternative imidic forms of Gln-51 were considered, for which such a H-bond is possible. Due to the high barrier of Gln-51 rotation a viable alternative is the Z-form of Gln-51 (see Fig. 4.4). The latter conserves the H-bond between its OH group and FMN-N5, which is also present in the Q51-E geometry.

QM/MM geometry optimizations (employing the very same QM/MM specifications as before) on conformations containing the E- and Z-forms of the Gln-51 moiety reveal that the E-form of Gln-51 is preferred over the Z-form, in spite of the additional H-bond. It appears that the Z-form of Gln-51 is destabilized by the neighboring Met-92 due to repulsion between the sulphur atom of Met-92 and the lone pair of the NH group of Gln-51.

At first glance it may seem as if the Q51-E structure contradicts experimental findings, since the expected H-bonding interaction is missing. There is however evidence for a fast transforming intermediate in AppA, BlrB[35] and Tll0078[190], which can be only trapped at low temperatures. The measured shift of the LE band in the UV/VIS spectrum for this intermediate amounts to about half of that measured for the final signaling state. Fukushima *et. al.*, for example, report about a 5nm red-shifted intermediate for the BLUF domain of Tll0078[190]. These authors conclude that further conformational changes in the protein environment in the vicinity of FMN are necessary to arrive at the signaling state with its characteristic 10nm red-shift feature. Similar conclusion was made by Jung *et. al.* for intermediates detected in AppA and BlrB BLUF domains[35].

Taking into account that biradical formation (via the proton transfers process driven by the CT state) and the subsequent *fast* biradical recombination steps studied in the present work involve movement of only light hydrogen atoms/protons and therefore are feasible in crystalline structures or low temperatures, we assign the Q51-E structure to the observed intermediate on the path towards the final signaling structure.

As mentioned in the introduction the position/orientation of the Trp-92 residue with respect to FMN is not well-defined in both dark and signaling

state conformations. One may therefore assume that this residue may well be involved in the aforementioned conformational changes upon receptor \leftrightarrow signaling state transitions. In the FT-IR study carried out by Masuda *et al.*[210] the mutation of the Trp-92 residue in AppA BLUF-domain was shown to weaken the hydrogen bonding network around FMN-O4, although the overall dark \rightarrow light red-shift for the C4=O4 IR-mode was not affected. Additionally, this mutation lead to a dramatic increase in the rate of thermal relaxation towards the dark state[210]. The presence of Trp-92 in BLUF domains therefore seems to stabilize the signaling state conformation. Hence it is of interest to compare the E- and Z-forms of the Gln-51 moiety for a Trp-in conformation.

Unfortunately, there is no crystal structure available for a possible Trp-in conformation of the BLUF-domain of BlrB. Based on Trp-in conformations published for AppA[49] and Slr1694[185] we made a model system for a possible Trp-in conformation for BlrB, which then was used as the starting structure for further QM/MM calculations (see section II. for details about the preparation of this model system). In contrast to the Met-in case, the Z-form of Gln-51 now is more stable than the E-form for the Trp-in model of the protein. This result now also provides an explanation for the much faster dark state recovery of the Trp-92 mutant, since the additional stabilization of the H-bond network around FMN-O4, caused by the latter residue, is missing in the mutant. The Trp-in conformation with Gln-51 in its (s-Z)-Z form is denoted as Q51-Z .

The transition from the Met-in Q51-E to the Trp-in Q51-Z geometry involves a further isomerization step (s-Z-E \rightarrow s-Z-Z cf. Fig 4.4, or E \rightarrow Z for short) of the Gln-51, which corresponds to inversion of the N-H bond. An alternative would be the thermal back reaction to the original dark state conformation via s-Z-E \rightarrow amide tautomerization. In order to estimate the barriers for the mentioned tautomerization and isomerization processes, gas-phase calculations (DFT/B3LYP employing the TZVP basis) for the Gln-51 model systems (as shown in Fig. 4.4) were performed. A barrier of about 33 kcal/mol for the imidic (s-Z-E) \rightarrow amine tautomerization was obtained whereas for the E \rightarrow Z isomerization step the barrier is about 28 kcal/mol. Additional water molecules, which enable a concerted proton transfer, reduce the barrier further to 22 kcal/mol as additional test calculations have shown.

Having three representative QM/MM optimized structures, dark, Q51-E , and Q51-Z in hand, the next step is to include the Asn-33 and Trp-92 residues in the QM-region. Experimentally, these have been shown to influence both UV/VIS spectra and IR-spectra as well as the signaling \rightarrow dark state transition rate, as mentioned earlier. Therefore they need to

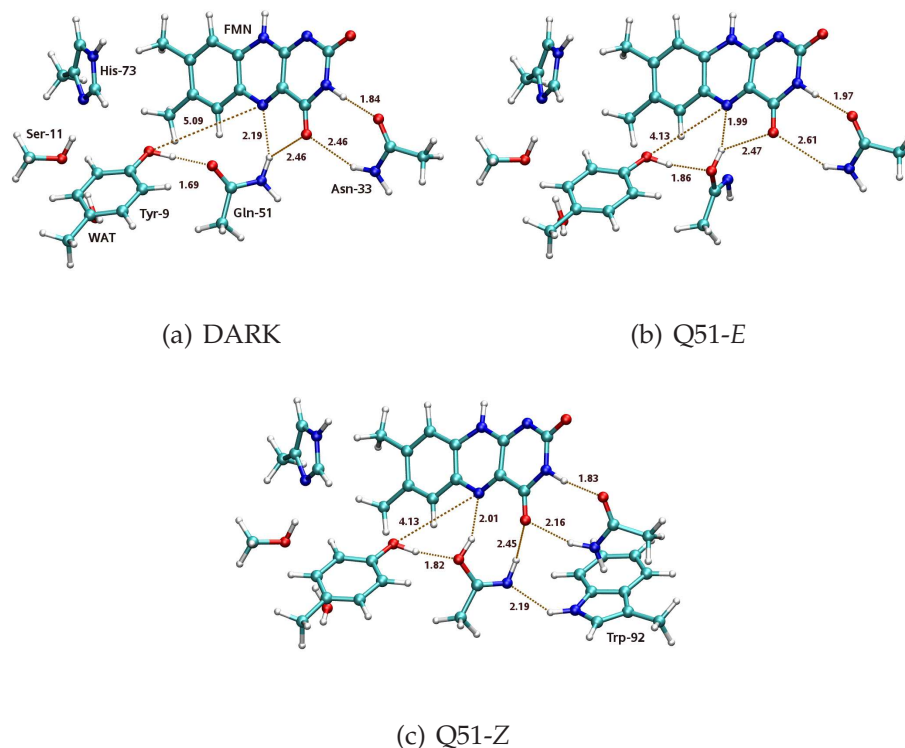


Figure 4.9: Optimized ground state QM/MM structures for the dark, Q51-E and Q51-Z conformations

be included at the more accurate QM level for the vertical excitation and vibrational frequency QM/MM calculations (see below). For the dark and Q51-E geometries the initial QM region (set A) was augmented by Asn-33 (set B). For the Q51-Z geometry set A was augmented by Asn-33 and Trp-92 (set C), respectively. The following observations, made by comparing the QM/MM optimized structures shown in Fig. 4.9, are complementary to the further QM/MM results discussed in subsections G. and H. which deal with the vertical excitation energy and vibrational frequency calculations performed for the mentioned three structures.

It is for example evident that the rather weak $\text{NH}_2\text{-N5}$ H-bond at the dark geometry, is replaced by a stronger OH-N5 H-bond (at the Q51-E geometry) after the direct biradical recombination process discussed above. On going from the Q51-E to the Q51-Z structure the length of this OH-N5 H-bond between FMN and Gln-51 remains virtually constant. Recalling that the electron density of the LE state increases considerably at N5 position of flavin (see Fig. 4.6(a)), this former observation is a clear indication that

the H-bond at N5 position, which we expect to be partially responsible for the slight red-shift of the band related to the LE state, as observed in the UV/VIS spectrum, is not disturbed by even the rather large conformational change of the Trp-Met flip. This is in agreement with the flash-photolysis experiment carried out by Gauden*et. al.*[211] where the authors paper conclude that the red-shifted absorption band of the signaling state appearing after less than 10ns after initial excitation of the dark structure, does not change thereafter within the observed time range of 10 ns to some ms.

Another strong H-bonding interaction between Asn-33 and FMN-O4 is formed during the signaling state formation as can be seen by comparing the dark and Q51-Z (our proposed dark and signaling state) conformations (Asn-33 H/FMN-O4 distance: 2.46Å vs. 2.16Å, see Fig. 4.9(a) and 4.9(c)). This is in line with crystal structures solved for BLUF domains corresponding to Trp-in conformations. Considering the slight increase in the electron density at the FMN-O4 position after excitation to the LE state, the stronger H-bond formed between FMN-O4 and Asn-33 could be of importance for inducing the red-shift observed in the UV/VIS absorption spectrum. In fact Asn-33 mutated BLUF domains show only half of the mentioned red-shift as that observed for the wild type[189] which is in agreement with the role proposed here for this residue.

This strengthened hydrogen bonding interaction between FMN-O4 and Asn-33 could also be the main reason for the red-shift in the IR vibrational C4=O4 stretch mode of FMN. There is a second H-bond between the Z-tautomer of Gln-51 and FMN-O4, only found after inversion of the N-H bond (discussed above), yet one expects the stronger H-bond (Asn-33/FMN-O4, Fig. 4.9(c)) to be *the* hydrogen bond formed during the signaling state formation hence causing the red-shift in the C4=O4 stretch mode.

4.3.7 Vertical excitation energies: comparison of dark, Q51-E and Q51-Z structures

Vertical excitation energies at the dark, Q51-E , and Q51-Z geometries were computed by applying both TD-DFT (using the B3LYP and BHLYP functionals) and CC2 response theory. It has been stated already above that we do not consider the TD-DFT (B3LYP) values as reliable, certainly not for the CT state, they are just given here for comparison. The QM/MM regions were specified as the sets A, B, and C, respectively and the results so obtained are compiled in Table4.3.

Locally excited state (FMN \rightarrow FMN*):

No substantial differences in the excitation energies of the LE state are apparent between the individual geometries. Red-shifts of 0.07 eV (9.8nm) and 0.1 eV (15.0nm) between the excitation energies of the dark and Q51-Z geometries were obtained at the level of CC2 response theory for the QM regions A and B/C, respectively. This is basically in agreement with experiment since all known BLUF domains do not feature any substantial spectral shift of the LE absorption band between dark and signaling state. The good agreement of our values above with the experimental red-shift of 10nm must be considered as fortuitous.

A shift of about 0.1 eV is a subtle feature and beyond the accuracy of the present approach, not so much because of the underlying quantum chemical method, i.e., CC2 response theory, but because of the MM environment based on far-reaching point charges. Test calculations, performed in the context of the present work, reveal, that modifications of charges quite far from the FMN chromophore still have an appreciable influence on the excitation energies. Moreover, the vertical excitation energies given in Table4.3 represent just one minimized configuration, whereas an ensemble average of excitation energies would be needed for a direct comparison with experiment (cf. e.g. Ref [212]).

Due to these reasons we cannot claim to really pick up the measured red-shift of 10nm with the methodology utilized here. However, our calculations clearly show a significant increase in the electron density at the N5 position on going from the ground to the LE excited state (see Fig.4.6).

Table 4.3: TD-DFT and CC2 response vertical excitation energies (in eV) calculated at QM/MM optimized dark, Q51-E, and Q51-Z structures. LE and CT represent the local excitation within FMN and the charge transfer excitation from Tyr-9 to FMN, respectively. Oscillator strengths are shown in parenthesis. QM set A is the default specification for the QM region of our QM/MM calculations. For the calculations involving the QM sets B (A+Asn-33) and C(A+Asn33+Trp92) the related ground state geometries were re-optimized accordingly.(see text)

		TDDFT-B3LYP		TDDFT-BHLYP		CC2	
Struc.	set	LE	CT	LE	CT	LE	CT
Dark	A	2.81(0.179)	1.860(0.000)	3.290(0.269)	3.170(0.001)	2.992	3.596
Q51-E	A	2.900(0.108)	1.833(0.000)	3.300(0.249)	3.240(0.000)	2.996	3.656
Q51-Z	A	2.699(0.037)	0.947(0.000)	3.175(0.176)	2.366(0.000)	2.923	2.732
Dark	B	2.872(0.100)	1.645(0.000)	3.338(0.286)	3.167(0.000)	3.031	3.627
Q51-E	B	2.900(0.108)	1.833(0.000)	3.356(0.295)	3.310(0.000)	3.047	3.695
Q51-Z	C	2.736(0.074)	1.105(0.000)	3.182(0.238)	2.528(0.000)	2.924	2.873

Based on this fact one may anticipate that the experimentally observed 10nm shift is related to changes in the H-bonds involving this atom of FMN.

Charge transfer state (Tyr-9 \rightarrow FMN):

The TD-DFT values exhibit a substantial drop in the energy of the CT state on going from the dark to the Q51-Z geometry. However, the CT state energies are considerably underestimated (particularly so for the B3LYP functional for reasons mentioned above) such that the gap between the LE and the CT states at the Q51-Z geometry becomes rather large. Furthermore, no conical intersection is predicted by TD-DFT. We conclude that TD-DFT alone is not able to provide a qualitatively correct picture of the photophysical behavior of the BLUF domain. CC2 response, on the other hand, certainly provides a qualitatively correct picture of the energetics of the low-lying LE and CT states in BLUF domains.

According to Table 4.3 CC2 response yields a decrease in the energy of the CT state by almost 0.9 eV on going from the dark to the Q51-Z geometry. For the dark geometry the CT state is energetically *above* the LE state with a gap of 0.6 eV, while for the Q51-Z geometry the CT state is 0.2 eV *below* the LE state. Evidently, the CT state is much more stable in the Q51-Z than in the dark geometry. It is tempting to assume that also the Frank-Condon point of the signaling state (Q51-Z geometry) is located close to a conical intersection seam between the LE and the CT surfaces and as a result populating the CT state via the LE state upon photo-excitation. Consequently, the CT state would be also instrumental for the photocycle initiated by the photo-excitation of the signaling state of the BLUF domain (*vide infra*).

Zirak *et. al.* have observed a lower fluorescence efficiency and a shortened fluorescence lifetime in the signaling state (compared with dark/receptor state) of BlrB[184], AppA[182] and Slr1694[188] BLUF domains. These data support our hypothesis of the above mentioned conical intersection near the Q51-Z geometry efficiently populating the CT state. Additionally, the ultra-fast transient absorption spectroscopy measurements by Gauden *et. al.*[209] indicate a much faster photocycle in the light-adapted form of the AppA BLUF domain. In fact, no FMN anionic semi-quinone (FMN $^{\bullet-}$) was observed in this work and the authors report on the formation of flavin neutral semi-quinone within some ps after photo-excitation, followed by ground state recovery (relaxation down to the signaling state conformation) after 60ps.

The Q51-Z geometry may explain a much faster photocycle by faster proton transfer steps due to stronger hydrogen bonds. The relatively weak H-bond between FMN-N5 and NH₂ of Gln-51 in the dark state is replaced

in the Q51-Z geometry by a stronger H-bond involving the hydroxy group of the Z-tautomer of Gln-51, which is more acidic, in the signaling state. Upon activation of the CT state in the Q51-Z conformation it is more probable that a concerted proton transfer from Tyr-9/OH to Gln-51/O ϵ and from Gln-51/OH to FMN-N5 takes place. This assumption can be justified by the fact that such a concerted proton transfer step, in contrast to the case of dark state, is mediated via a single atom, namely the O ϵ of Gln-51.

One may assume further that such a concerted motion would occur even faster than the P₁+P₂ proton transfer steps discussed for the dark structure, hence explaining the difficulty Gauden*et. al.* had observing the FMN anionic semi-quinone (FMN \bullet^-)[209]. After this biradical formation the NH group of Gln-51 is still within in a (somewhat weak) hydrogen bonding network involving the FMN-O4 and Trp-92. The orientation of the NH group prevents it from acting as a H acceptor, since the lone pair of the Gln-nitrogen (N ϵ) atom, pointing towards Trp-92, is already involved in a H-bonding interaction with the latter. For the radical recombination process to occur, the hydrogen atom at N5 is transferred back to Tyr-9 OH (again) via a possibly concerted hydrogen atom transfer in the opposite direction to that inducing the biradical in the first place, i.e. FMN-N5 \rightarrow Gln-51 O ϵ and Gln-51 O ϵ \rightarrow Tyr-9 OH. This last mentioned mechanism is probably more compatible with experimental observations since other biradical recombination paths may well involve larger conformational changes (on a longer timescale) such as the rotation of Gln-51 or the dislocation of Tyr-9, etc.

There is a further charge transfer state, namely Trp-92 \rightarrow FMN, which is of some importance. At the Q51-Z geometry where the distance between FMN and Trp-92 is substantially decreased relative to the Met-in dark state structure it constitutes the lowest electronically excited state with a CC2 vertical excitation energy of 2.618 eV. The corresponding value for the Tyr-9 \rightarrow FMN CT state amounts to 2.873 eV (cf. Table 4.3). TD-DFT excitation energies for the Trp-92 \rightarrow FMN CT state amount to 2.17 eV (BHLYP) and 0.84 (B3LYP), again significantly underestimating the CC2 value (in particular for the latter functional). The oscillator strength of this CT state, calculated both at the level of TD-DFT and CC2, is essentially zero. Consequently, direct population via photo-excitation is not possible and the LE state has to be involved (possibly via conical intersection). Since the Tyr-9 \rightarrow FMN charge transfer state is energetically closer to the LE state (actually rather close, cf. Table 4.3), one can anticipate that the Tyr-9 moiety still is the primary electron donor in the signaling state of BLUF domains and the Trp-92 most probably is playing a secondary role. Interestingly, Gauden*et. al.*[183] suggested the possibility of a competition between the two CT pro-

cesses discussed here. The photo cycle of BLUF domains in the signaling state is subject of further research in our group.

4.3.8 QM/MM vibrational frequency calculation

Apart from the calculations of vertical excitation energies discussed above, also QM/MM harmonic vibrational frequencies in the electronic ground state were computed at the three geometries dark, Q51-*E*, and Q51-*Z*, respectively. The resulting values for the C4=O4 stretch mode are given in Table 4.4. The biradical recombination process outlined in our study leads to the Q51-*E* geometry (Gln-51 in imidic form), for which no substantial strengthening of the H-bond network involving FMN-O4 is taking place. The calculated shift of the FMN C4=O4 stretch mode relative to the dark geometry therefore is rather small (cf. Table 4.4).

Table 4.4: Stretch and bend mode frequencies (in cm^{-1}) for the dark, Q51-*E* and Q51-*Z* structures calculated at the B3LYP/MM level of theory. ^a the C4=O4 and C2=O2 stretch modes couple rather strongly with the N3-H3 mode. ^b the C=O mode in Gln-51 couples strongly with the NH2 bend coordinate at the dark state geometry.

Struc.	FMN modes				Gln-51 modes	
(set)	C4=O4 ^a	C2=O2 ^a	N1-C10	C4a-N5	C=O ^b /C=N	NH2-bend
dark (B)	1797.8	1745.3	1590.7	1656.9	1716.3	1653.4
Q51- <i>E</i> (B)	1795.2	1736.7	1600.4	1662.6	1760.3	-
Q51- <i>Z</i> (C)	1791.0	1737.3	1597.9	1656.3	1773.4	-

At the first sight this appears to be in contradiction with the experiments measuring a red-shift of about 10 cm^{-1} of the FMN C4=O4 stretch mode on going from the dark to the signaling state of the BLUF domain. However, a recent FT-IR study by Stellinget *al.* on AppA BLUF domains[46] reveals that this shift does not appear in the first 2ns after photo-excitation of the dark-adapted structure, whereas the biradical recombination is reported to be completed within 2-3ns in the AppA[211] and Slr1694[47] BLUF domains.

This implies that the strengthening of the H-bond network involving FMN-O4 at the O4 position and the formation of the intermediate arrived at after biradical recombination (the Q51-*E* geometry proposed here) are two separate steps occurring at different times in the photo-cycles, i.e., the strengthening of the H-bond network occurs *later* than the formation of

the Q51-E geometry. Yet this is in excellent agreement with our frequency calculations, where the final Q51-Z geometry, i.e., our postulated signaling state geometry, exhibits basically the experimentally observed shift (7cm^{-1} vs. 10cm^{-1} reported in Ref. [46]). We note in passing that for the alternative mechanism involving rotation of the Gln-51 moiety a much larger shift of 40cm^{-1} was computed for the corresponding final signaling state[48]. Further stretch and bend modes calculated in the context of present work are shown in table Table4.4. The C2=O2 stretch mode, also showing a red shift upon transition to the signaling state, is lower than that of C4=O4 which is in line with experimental observations[181, 46].

A further agreement with experiment is the rather high contribution of the FMN N3-H3 stretch coordinates to both of these carbonyl modes[181, 46]. The C=O and C=N modes of Gln-51, also shown in Table4.4, include contributions from the NH2 bend coordinates of this residue. The mode found at 1653 cm^{-1} is assigned to the bend mode of Gln-51 which also couples with the C=O stretch coordinate of the latter residue. This mode is however not present in the Q51-E and Q51-Z structures, where the Gln-51 side chain is in its imidic form. More accurate QM/MM calculations at a higher level of theory (e.g. MP2) are necessary for a more robust assignment of these modes to the experimental ones.

4.4 Conclusions

Extensive QM/MM calculations were performed, applying also high-level *ab initio* methods like Coupled Cluster (CC2) response theory in order to lay out the reaction path leading from the dark state to the signaling state structure after initial photoexcitation of the former. A Tyr-9 \rightarrow FMN charge transfer state is populated via conical intersection after triggering an electronically local excited state of the FMN subunit. The latter has, in contrast to the charge transfer state, a rather large oscillator strength. Populating charge transfer states, which usually have only low oscillator strength, via conical intersection appears to be a common way to achieve photoinduced charge transfer, at least for systems based on flavin.

The reaction cascade then proceeds downhill on the charge transfer surface, inducing two proton transfer steps, and finally leads via a second conical intersection to a non-zwitterionic ground state biradical. These two proton transfer steps transform the Gln-51 moiety, the key player in this whole mechanism, to its imidic tautomer. Biradical recombination, a further step in the cascade, then transforms this tautomer to the E-isomer (E-form). A subsequent second isomerization, triggered by the exchange of the Met-94

and Trp-92 residues in its vicinity further stabilizes the imidic form of Gln-51 (Z-form). Accompanied by this *flip* is a spatial dislocation of the Asn-33 residue, moving closer to the FMN chromophore. It turns out to be this effect that is mainly responsible for the experimentally observed red-shift in the flavin C4=O4 stretch frequency.

The thermal back-reaction from the long living signaling state involves the back-tautomerization of Gln-51. The related barrier is estimated to be rather high, particularly so, if the imidic tautomer is stabilized in the Z-form after the interchange of Met-94 and Trp-92. We assume that this Met-94 vs. Trp-92 flip actually is the trigger for the large conformational change of the protein transmitting the signal to the biological environment. In its dark state and under aerobic conditions, AppA binds to Purple bacteria photo system Regulator (PpsR), a transcription factor, hence forming a complex which is not able to bind to DNA and therefore initiates gene expression. The oligomerisation state of AppA changes in its light-induced signaling state[213] but the mechanism of PpsR release, leading to gene expression inhibition, is not fully known. The proposed mechanism also provides a hint for the role of the strictly conserved Met-94 residue next to Gln-51. Methionine is the only amino acid which can only act as an H-acceptor, hence it forms an attractive interaction with Gln-NH₂ but is repulsive towards the proposed Gln-Z isomer.

None of the reaction steps presented here do involve any rotation of the Gln-51 sidechain. Rotation of Gln-51 was previously postulated by others to play a key role in the mechanism, but the present study shows that the barriers of rotation are considerably higher than those of direct proton transfer as conjectured here. Interestingly, the mechanism implies only photoisomerization of the protein environment but not of the FMN itself. The FMN chromophore just plays the role of an antenna and mediator for the tautomerization and subsequent isomerization of Gln-51. This, to our knowledge, is unique for biological photoreceptors.

Chapter 5

2'-Oxoethyl Flavin Revisited

One of the projects of the Graduate College, pursued by Prof. B. König (Department of Organic Chemistry, University of Regensburg) and his coworker Dr. J. Svoboda, was to find an alternative to the rather complicated and tedious Kuhn-Synthesis[214] of flavin and flavin-based molecules. The main idea is to use commercially available riboflavin (**1**) and use it as a starting material. An intermediate in this alternative synthesis, formed after oxidative degradation of riboflavin, is the 2'-Oxoethyl flavin (**2**) (see Fig. 5.1)

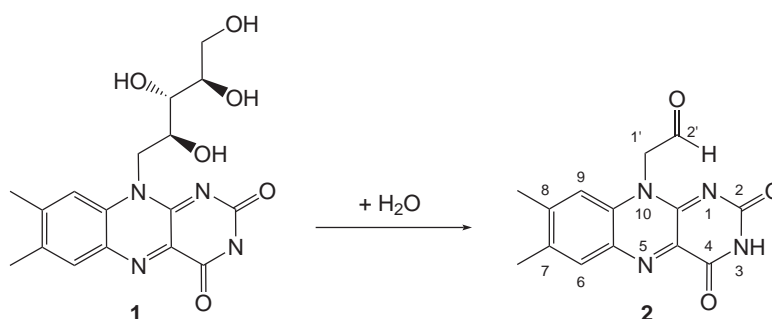


Figure 5.1: Oxidative degradation of riboflavin (**1**) to 2'-oxoethyl flavin (**2**) and numbering scheme; conditions: periodic acid, sulphuric acid, water, r. t., 49 %. For more experimental detail see Ref.[215]

It turned out however that the hydrated *gem*-diol form (**4**) is predominantly preferred over the 2-Oxoethyl flavin (**2**) (see Fig. 5.2). This behaviour is a rather unusual for an aldehyde.

For comparison, a similar experiment was carried out with the structurally related 2-phenylpropionic aldehyde **5** (Fig. 5.3). Again, an intensity increase of the *gem*-diol **6** signals at the expense of the signals for aldehyde **5**

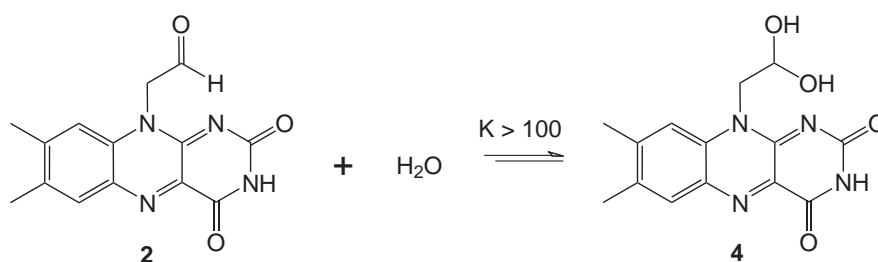


Figure 5.2: Equilibrium hydration of 2'-oxoethyl flavin (2).

was observed, but in this case, the mixture equilibrated at 45:55 (*gem*-diol 6:aldehyde 5) ratio which corresponds to an equilibrium constant of 0.8, as expected for aldehydes in general[215].

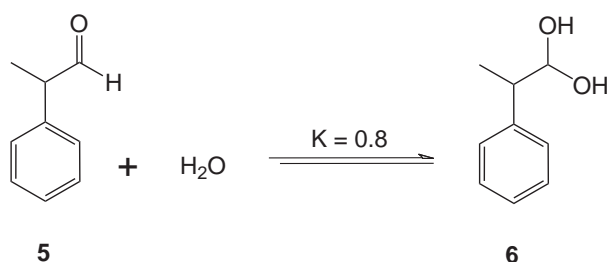


Figure 5.3: Equilibrium hydration of 2-phenylpropionic aldehyde (5).

In collaboration with the group of Prof. König, we have performed *local* MP2 (LMP2) calculations for the molecules involved in the hydration reactions shown in Fig. 5.2 and Fig. 5.3.

The aim of these calculations was to understand the rather unusual behaviour of 2'-oxoethyl flavin (2). Of special interest here is to estimate the strength of the hydrogen bond formed between the hydroxy group and the N1 atom of flavin in the *gem*-diol (4) molecule. Such a H-bonding interaction (missing in the *gem*-diol (6) system) is expected to be main cause for the stability of the *gem*-diol (4) compared with the 2'-oxoethyl flavin (2).

Geometry optimisations of 2'-oxoethyl flavin (2) and its *gem*-diol form were carried out with the efficient analytic LMP2 energy gradient method[199] using the aug-cc-PVDZ AO basis set of Dunning [202]. Single point energy calculations at these geometries were performed employing the more extended aug-cc-PVTZ and aug-cc-PVQZ sets, respectively, which were used to extrapolate the correlation energy at the basis set limit (two-point extrapolation formula, Ref. [216]).

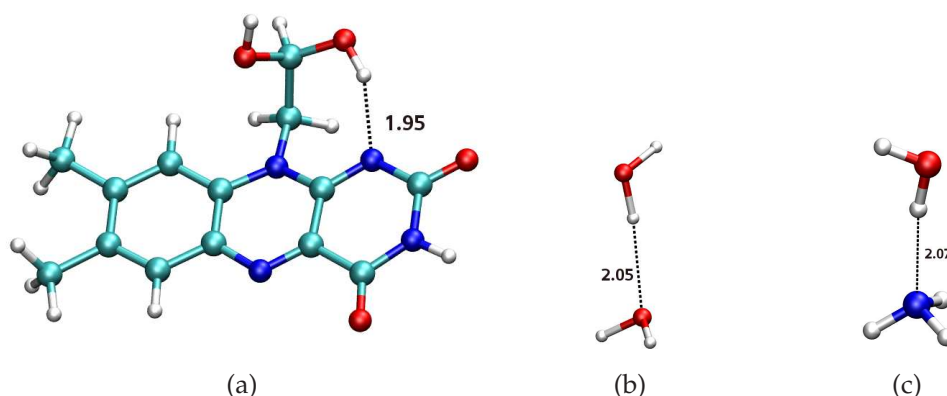


Figure 5.4: The hydrogen bond at the N1 position of flavin(a) compared with that found in water dimer (b) and water-ammonia (b) complexes as obtained by performing LMP2 calculation with the aug-cc-pVDZ basis set (see text for further details)

Analogous calculations were also performed for 2-phenylpropionic aldehyde and its *gem*-diol form to have, as a reference system, an aldehyde with “normal” chemical behaviour¹

For the *gem*-diol form **4** of 2'-oxoethyl flavin molecule the calculations predict the formation of an intramolecular hydrogen bond between one of the hydroxy groups of the diol and the nitrogen atom in position 1 of the flavin skeleton. The length of this hydrogen bond is comparatively short, *i.e.*, 1.96 Å *vs.* 2.05 and 2.07 Å for the water dimer and the water ammonia complex, respectively, calculated at the same level of theory (Fig. 5.4).

In order to assess the strength of this hydrogen bond additional *constrained geometry* optimisations were performed for a sequence of different C-1'-C-2'-OH dihedral angles (see Fig. 5.5). A barrier height of 8.34 kcal/mol at

¹The *ab initio* calculations were performed with the local MP2 method as implemented in the MOLPRO [173] program package, employing the density fitting approximation for the electron repulsion integrals [217, 199]. The augmented correlation consistent AO basis sets aug-cc-pVXZ of Dunning [174, 202] were used (X = D for geometry optimizations, X = T,Q for single point energies), along with the related fitting basis sets optimized for DF-MP2 [175].

For the Hartree-Fock energy and the related component of the LMP2 gradient the JK-fitting basis sets of Weigend [101] related to the cc-pV(X+1)Z AO basis, respectively, were employed. Local orbitals were generated according to the Pipek-Mezey localization scheme [218]. Pair domains were constructed with the Boughton-Pulay procedure [219] using a completeness criterion of 0.98. The BP domains then were extended by all next nearest neighbour centres. The occupied orbital pair list remained un-truncated in all calculations.

The density functional calculations were carried out by using the TURBOMOLE application package [220].

the b

e.

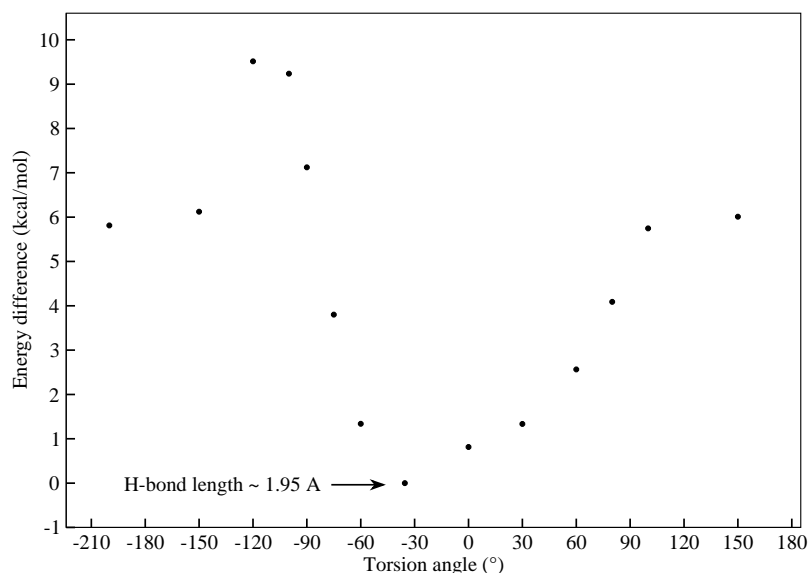


Figure 5.5: The LMP2 energy profile along the torsion angle C-1'-C-2'-OH in *gem*-diol which leads to the breaking of the hydrogen bond formed at the N1 position.

Due to the absence of sterical hindrance this barrier height appears to be a reasonable estimate for the strength of the intramolecular hydrogen bond, which is substantially stronger than the hydrogen bond of the water dimer (4.94 kcal/mol) or even of the water-ammonia dimer (6.48 kcal/mol). For the electronic contribution to the hydration reaction energy of 2'-oxoethyl flavin (**2**) a value of -47.2 kcal/mol (extrapolated to the basis set limit, -46.8 kcal/mol for the aug-cc-pVQZ basis alone) was obtained. This is -5.5 kcal/mol more than for the reference system, again reflecting the enhanced stability of the former due to intramolecular hydrogen bond formation. Based on these electronic reaction energies, the related free energy differences at room temperature were assessed using harmonic vibrational frequencies calculated at the level of density functional theory (B3-LYP hybrid functional, TZVP basis set [221]).

The free energy differences for the hydration reactions so obtained amount to -38.2 kcal/mol and -26.9 kcal/mol for 2'-oxoethyl flavin (**2**) and 2-phenylpropionic aldehyde (**5**), respectively. Due to the underlying approximation of an ideal solution these two values certainly have to be taken with care. However, the error imposed by this model is likely to cancel to a large extent in the difference between these two free energy differences, which amounts to -11.3 kcal/mol.

Thus we can infer from these free energy calculations that (i) the free reaction energies are smaller (absolute value) than the corresponding pure electronic reaction energies, and (ii) zero-point energy corrections and finite temperature entropic effects disfavour the *gem*-diol form to lesser extent for 2'-oxoethyl flavin (**2**) than for 2-phenylpropionic aldehyde (**5**) as the reference system (see Fig. 5.6).

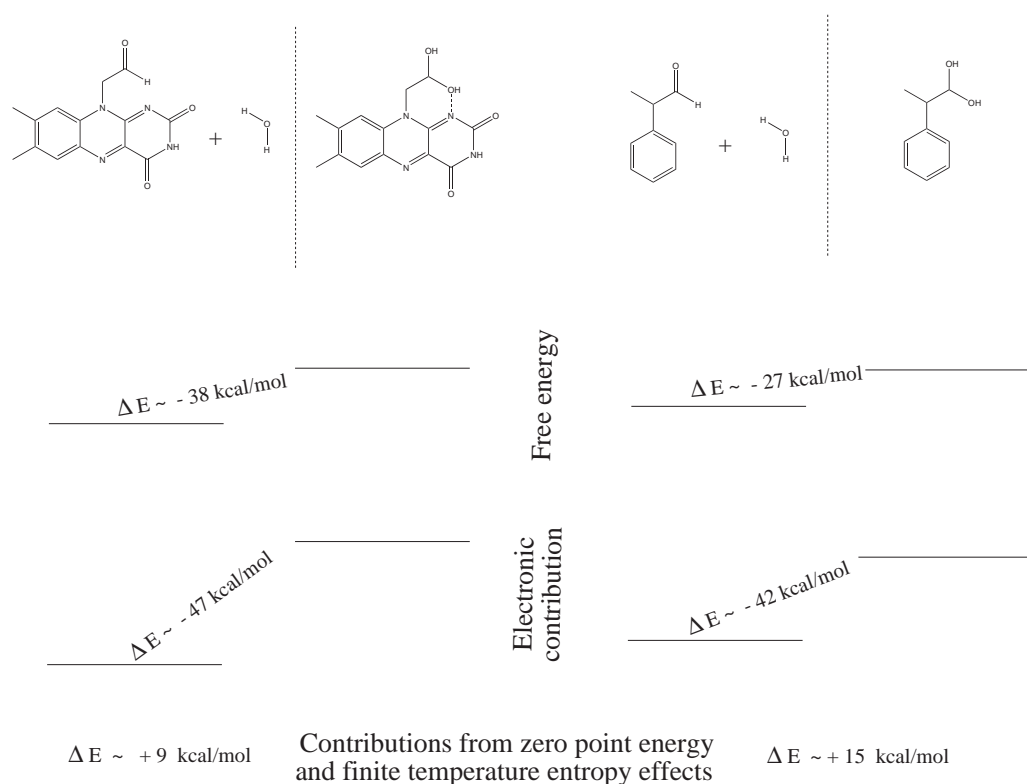


Figure 5.6: Comparison of the electronic and free energy contributions to the hydration calculated at the LMP2 level for both flavin and 2-phenylpropionic systems.

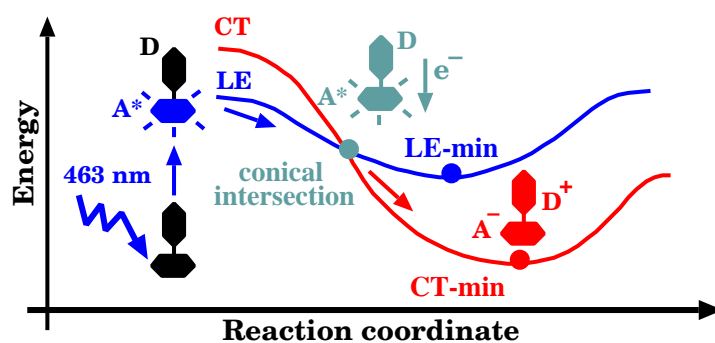
To summarise, we conclude from our calculations and experimental findings, that the intramolecular hydrogen bond occurring in the *gem*-diol form **4** of 2'-oxoethyl flavin (**2**) leads to a stabilisation of the diol over the aldehyde to such an extent that the aldehyde form **2** can barely be observed by the spectrometric methods applied in this work.

Chapter 6

Summary

In this project an artificial photoreceptor of cryptochrome type, phenothiazine-pyrene-flavin dyad (*PF-dyad*), and a family of biological photoreceptors, the *BLUF* domain proteins, were studied using *ab initio* calculations. The covalently unbound flavin chromophore, in its oxidized form, is sensitive to the blue and near UV part of solar spectrum and plays a key role in the photochemistry and photophysics of both systems.

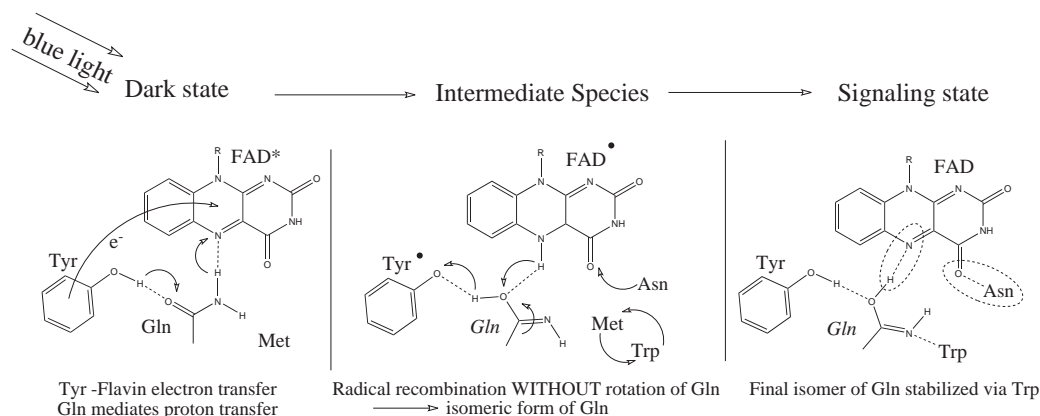
Experimental data for the *PF-dyad* indicate both low fluorescence quantum yield and formation of charged-separated (phenothiazine⁺-flavin⁻) species upon photoexcitation. These findings seem to be connected as photo-induced charge transfer processes are known to cause fluorescence quenching in donor-acceptor system. In the case of the *PF-dyad*, the initial excitation leads to the population of the locally excited state (LE) of flavin which has a high oscillator strength. The charge transfer (CT) state however has a low oscillator strength and therefore not directly activated via photon absorption. Our calculations show that a conical intersection between the two LE and CT states is indeed responsible for the activation of the latter, hence explaining the above mentioned experimental findings.



A series of constrained geometry optimizations on the CT state surface of the *PF-dyad* have been carried out at the TD-DFT level of theory. It is well documented that TD-DFT yields erroneous and untrustworthy results for the excitation energies of CT states, hence a more reliable method, namely CC2, was applied to correct for this shortcoming.

The environment of a donor-acceptor system may have strong influence on the energetic position of the conical intersection seam. Depending on the nature of the solvent which surrounds the molecule and those states which are involved in its photochemistry, the lowest energy region of the conical intersection seam can either be stabilized, as observed in our case study (solvent = acetonitrile), or even destabilized.

For a correct description of photo-induced processes therefore it is crucial to include the environment explicitly in the calculations. A popular method used these days is the hybrid QM/MM approach, where the active center of a (photo-)chemical reaction is treated with laws of quantum mechanics, while the remaining part of the system is described via molecular mechanics. Such an approach was used in this project to study the signaling state formation in the BLUF domain family of biological photoreceptors. Having all the experimentally known data on BLUF domains in mind, we tried to answer some of the open questions by setting up and performing QM/MM calculations.



The main finding of this QM/MM study is that in contrast to all other families of known biological photoreceptors, where the chromophore either undergoes isomerization (e.g. rhodopsin) or photo-adduct formation (e.g. LOV domain), in *BLUF* domains a flavin-neighboring and highly-conserved amino acid, glutamine, is in fact isomerized into its imidic form. The imidic isomer is further stabilized by the interchange of the two me-

thionine and tryptophane residues within the flavin binding pocket. The transition to the dark state structure (i.e. dark state recovery) can only occur thermally since the protein is locked in its light-induced state. This stable and long-living conformation opens the way for the subsequent signal transduction process.

Photo-induced electron transfer is an important part of the mechanisms studied in this project. Cryptochrome and BLUF proteins, belonging to the naturally found photoreceptors, make use of this process in their signaling state formation cascade and the flavin chromophore, well known for its catalytic functions, plays the key role. Nature has found ways to enhance such capabilities by "tuning" the environment of the flavin in such a way that (i) an electron donor, e.g. the highly conserved tyrosine in the case of BLUF domains, is placed in the vicinity of the chromophore (electron acceptor), and (ii) the low energy region of the conical intersection between the LE and CT states is energetically accessible. Once these two conditions are fulfilled, the photo-induced charge transfer process initiates a whole cascade of events which finally lead to the desired cellular response.

Bibliography

- [1] Montgomery, B. L. *Mol. Microb.* **2007**, 64, 16.
- [2] Purcell, E.; Crosson, S. *Curr. Opin. Microb.* **2008**, 11, 1.
- [3] van der Horst, M. A.; Hellingwerf, K. J. *Acc. Chem. Res.* **2004**, 37, 13.
- [4] Hoff, W. D.; Jung, K. H.; Spudlich, J. L. *Annu. Rev. Biophys. Biomol. Struct.* **1997**, 26, 223.
- [5] Quail, P. H. *Philos. Trans. R. Soc. London B: Biol. Sci.* **1998**, 353, 1399.
- [6] Kort, R.; Hoff, W. D.; van West, M.; Kroon, A. R.; Hoffer, S. M.; Vlieg, K. H.; Crielaand, W.; Beeumen, J. J. V.; Hellingwerf, K. J. *EMBO J.* **1996**, 15, 3209.
- [7] Huala, E.; Oeller, P. W.; Liscum, E.; Han, I. S.; Larsen, E.; Briggs, W. R. *Science* **1997**, 278, 2120.
- [8] Swartz, T. E.; Corchnoy, S. B.; Christie, J. M.; Lewis, J. W.; Szundi, I.; Briggs, W. R.; Bogomolni, R. A. *J. Biol. Chem.* **2001**, 276, 36493.
- [9] Ahmad, M.; Cashmore, A. R. *Nature* **1993**, 366, 162.
- [10] Gomelsky, M.; Klug, G. *Trends Biochem. Sci.* **2002**, 27, 497.
- [11] Hitomi, K.; Okamoto, K.; Daiyasu, H.; Miyashita, H.; Iwai, S.; Toh, H.; Ishiura, M.; Todo, T. *Nuc. Acid. Res.* **2000**, 28, 2353.
- [12] Huang, Y.; Baxter, R.; Smith, B.; Partch, C.; Colbert, C.; J., J. D. *Proc. Natl. Acad. Sci.* **2006**, 103, 17701.
- [13] Mees, A.; Klar, T.; Gnau, P.; Hennecke, U.; Eker, A.; Carell, T.; Essen, L.-O. *Science* **2004**, 306, 1789.
- [14] Lukin, M.; de Los Santos, C. *Chem. Rev.* **2006**, 106, 607.

- [15] Behrens, C.; Cichon, M.; Grolle, F.; Hennecke, U.; Carell, T. *Top. Curr. Chem.* **2004**, *236*, 187.
- [16] Selby, C.; Sancar, A. *Proc. Natl. Acad. Sci.* **2006**, *103*, 17696.
- [17] Sancar, A.; Sancar, G. *Annu. Rev. Biochem.* **1987**, *57*, 29.
- [18] Ruprecht, C.; Goodgal, S.; Herriot, R. M. *J. Gen. Physiol.* **1958**, *41*, 451.
- [19] Sancar, A. *Chem. Rev.* **2003**, *103*, 2203.
- [20] Sancar, G.; Smith, F.; Reid, R.; Payne, G.; Levy, M.; Sancar, A. *J. Biol. Chem.* **1987**, *262*, 478.
- [21] Payne, G.; Sancar, A. *Biochemistry* **1990**, *29*, 7715.
- [22] Shen, Z.; Strauss, J.; Daub, J. *Chem. Commun.* **2002**, 460.
- [23] Winnik, F. M. *Chem. Rev.* **1993**, *93*, 587.
- [24] Shirdel, J.; Penzkofer, A.; Procházka, R.; Shen, Z.; Daub, J. *Chem. Phys.* **2007**, *336*, 1.
- [25] Shirdel, J.; Penzkofer, A.; Procházka, R.; Shen, Z.; Daub, J. *Chem. Phys.* **2007**, *337*, 99.
- [26] Shirdel, J.; Penzkofer, A.; Procházka, R.; Shen, Z.; Strauss, J.; Daub, J. *Chem. Phys.* **2007**, *331*, 427.
- [27] Shirdel, J.; Penzkofer, A.; Procházka, R.; Daub, J.; Hochmuth, E.; Deutzmann, R. *Chem. Phys.* **2006**, *326*, 489.
- [28] Daub, J.; Engl, R.; Kurzawa, J.; Miller, S. E.; Schneider, S.; Stockmann, A.; Wasielewski, M. R. *J. Phys. Chem. A* **2001**, *105*, 5655.
- [29] Schneider, S.; Kurzawa, J.; Stockmann, A.; Engl, R.; Daub, J.; Matousek, P.; Towrie, M. *Chem. Phys. Lett.* **2001**, *348*, 277.
- [30] Shen, Z.; Procházka, R.; Daub, J.; Fritz, N.; Acar, N.; Schneider, S. *Phys. Chem. Chem. Phys.* **2003**, *5*, 3257.
- [31] Acar, N.; Kurzawa, J.; Fritz, N.; Stockmann, A.; Roman, C.; Schneider, S.; Clark, T. *J. Phys. Chem. A* **2003**, *107*, 9530.
- [32] Stockmann, A.; J.Kurzawa; Fritz, N.; Acar, N.; Schneider, S.; Daub, J.; Engl, E.; Clark, T. *J. Phys. Chem. A* **2002**, *106*, 7958.

- [33] Gilbert, A.; Baggott, J. . In *Essentials of Molecular Photochemistry*; Blackwell Scientific Publications: Oxford, 1991.
- [34] Jung, A.; Domratcheva, T.; tarutina, M.; Wu, Q.; ko, W.-H.; Shoeman, R.; Gomelsky, M.; Gardner, K. H.; Schlichting, I. *Proc. Natl. Acad. Sci.* **2005**, *102*, 12350.
- [35] Jung, A.; Reinstein, J.; Domratcheva, T.; Shoeman, R. L.; Schlichting, I. *J. Mol. Biol.* **2006**, *362*, 717.
- [36] Gomelsky, M.; Kaplan, S. *J. Bacteriol.* **1995**, *177*, 4609.
- [37] Gomelsky, M.; Kaplan, S. *J. Bacteriol.* **1997**, *179*, 128.
- [38] Gomelsky, M.; Kaplan, S. *J. Biol. Chem.* **1998**, *177*, 35319.
- [39] Masuda, S.; Bauer, C. *Cell* **2002**, *110*, 613.
- [40] Iseki, M.; Matsunaga, S.; Murakami, A.; Ohno, K.; Shiga, K.; Yoshida, K.; Sugai, M.; Takahashi, T.; Hori, T.; Watanabe, M. *Nature* **2002**, *415*, 1047.
- [41] Crosson, S.; Moffat, K. *Plant Cell* **2002**, *14*, 1067.
- [42] Kottke, T.; Hegemann, P.; Dick, B.; Heberle, J. *Biopolymers* **2006**, *82*, 373.
- [43] Christie, J. *Annu. Rev. Plant Biol.* **2007**, *58*, 21.
- [44] Nakamura, H.; Okajima, K.; Ikeuchi, M.; Noguchi, T. *J. Am. Chem. Soc.* **2008**, *130*, 12884.
- [45] Sadeghian, K.; Bocola, M.; Schütz, M. *J. Am. Chem. Soc.* **2008**, *130*, 12501.
- [46] Stelling, A. L.; Ronayne, K.; Nappa, J.; Tonge, P.; Meech, S. *J. Am. Chem. Soc.* **2007**, *129*, 15556.
- [47] Gauden, M.; van Stokkum, I. H. M.; Key, J. M.; Lührs, D. C.; van Grondelle, R.; Hegemann, P.; Kennis, J. T. M. *Proc. Natl. Acad. Sci.* **2006**, *103*, 10895.
- [48] Domratcheva, T.; Grigorenko, B.; Schlichting, I.; Nemukhin, A. V. *Biophys. J* **2008**, *94*, 3872.

- [49] Anderson, S.; Dragnea, V.; Masuda, S.; Ybe, J.; Moffat, K.; Bauer, C. *Biochemistry* **2005**, *44*, 7998.
- [50] Jensen, F. . In *Introduction to computational chemistry*; J. Wiley: New York, 1998.
- [51] Yarkony, D. R. *Rev. Mod. Phys.* **1996**, *68*, 985.
- [52] Sicilia, F.; Blancafort, L.; Bearpark, M.; Robb, M. *J. Chem. Theor. Comput.* **2008**, *4*, 257.
- [53] Pulay, P. Analytical Derivative Techniques and the Calculation of Vibrational Spectra. In *Modern Electronic Structure Theory*, Vol. 2; Yarkony, D. R., Ed.; World Scientific: 1995.
- [54] MacKerell, A. J. *Comput. Chem.* **2004**, *25*, 1584.
- [55] van Gunsteren, W.; Berendsen, H. *Angew. Chem. Int. Ed. Eng.* **1990**, *29*, 992.
- [56] Ryckaert, J.; Ciccotti, G.; Berendsen, H. *J. Comput. Phys.* **1977**, *23*, 327.
- [57] Kirkpatrick, S.; Jr, C. G.; Vecchi, M. *Science* **1983**, *220*, 671.
- [58] Warshel, A.; Levitt, M. **1976**, *103*, 227.
- [59] Senn, H. M.; Thiel, W. *Top. Curr. Chem* **2007**, *268*, 173.
- [60] Sherwood, P. Hybrid quantum mechanical/molecular mechanics approaches. In *Modern Methods and Algorithms of Quantum Chemistry*; Grotendorst, J., Ed.; NIC Series 1; John von Neumann Institute for Computing (NIC): 2000.
- [61] Murphy, R. B.; Philipp, D. M.; Friesner, R. A. *J. Comput. Chem.* **2000**, *21*, 1442.
- [62] Dittrich, M.; Yu, J.; Schulten, K. *Top. Curr. Chem.* **2007**, *268*, 319.
- [63] Lin, H.; Truhlar, D. G. *Theor. Chem. Acc.* **2007**, *117*, 185.
- [64] Sieber, S.; Humbel, S.; Morokuma, K. *J. Chem. Phys.* **1996**, *105*, 1959.
- [65] Svensson, M.; Humbel, S.; Froese, R.; Matsubara, T.; Sieber, S.; Morokuma, K. *J. Phys. Chem* **1996**, *100*, 19357.

- [66] MacKerell, A. D. *et al.* *J. Phys. Chem. B.* **1998**, *102*, 3586.
- [67] Pavelites, J. J.; Gao, J. L.; Bash, P. A.; MacKerell, A. D. *J. Comput. Chem.* **1997**, *18*, 221.
- [68] Foloppe, N.; MacKerell, A. D. *J. Comput. Chem.* **2000**, *21*, 86.
- [69] MacKerell, A. D. *J. et al.* *FASEB J* **1992**, *6A*, 143.
- [70] Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.
- [71] Smith, W.; Yong, C.; Rodger, P. *Molecular Simulations* **2002**, *28*, 385.
- [72] Sherwood, P. *et al.* *J. Mol. Struct. (Theochem.)* **2003**, *1*, 632.
- [73] www.chemshell.org.
- [74] www.tcl.tk.
- [75] Baker, J.; Kessi, A.; Delley, B. *J. Chem. Phys.* **1996**, *105*, 192.
- [76] Singh, U. C.; Kollman, P. *J. Comput. Chem.* **1986**, *7*, 718.
- [77] Parr, R.; Yang, W. . In *Density Functional Theory of Atoms and Molecules*; Oxford University Press: Oxford, 1989.
- [78] Helgaker, T.; Jørgensen, P.; Olsen, J. . In *Electronic Structure Theory*; J. Wiley: New York, 2000.
- [79] Tsuzuki, S.; Lüthi, H. *J. Chem. Phys.* **2001**, *114*, 3949.
- [80] Zhao, Y.; Truhlar, D. *Acc. Chem. Res.* **2008**, *41*, 157.
- [81] Grimme, S. *J. Comput. Chem.* **2004**, *12*, 1463.
- [82] Mogado, C.; Vincent, M.; Hillier, I.; Shan, X. *Phys. Chem. Chem. Phys.* **2007**, *9*, 448.
- [83] Černý, J.; Hobza, P. *Phys. Chem. Chem. Phys.* **2007**, *9*, 5291.
- [84] Bruenker, R.; Peyerimhoff, S. *Mol. Phys.* **1978**, *35*, 771.
- [85] Roos, B.; Andersson, K.; Flüscher, M. *Chem. Phys. Lett.* **1992**, *192*, 5.
- [86] Laidig, W.; Bartlett, R. *Chem. Phys. Lett.* **1984**, *104*, 424.
- [87] Celani, P.; Werner, H.-J. *J. Chem. Phys.* **2000**, *112*, 5546.

- [88] Hill, J.; Platts, J.; Werner, H.-J. *Phys. Chem. Chem. Phys.* **2006**, *8*, 4072.
- [89] Grimme, S. *J. Chem. Phys.* **2003**, *118*, 9095.
- [90] Bachorz, R. A.; Bischoff, F.; Höfener, S.; Klopper, W.; Ottiger, P.; Leist, R.; Frey, J. A.; Leutwyler, S. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2758.
- [91] Grimme, S. *Chem Eur. J.* **2004**, *10*, 3423.
- [92] Saebø, S.; Pulay, P. *Annu. Rev. Phys. Chem.* **1993**, *44*, 213.
- [93] Knowles, P.; M.Schütz,; Werner, H.-J. Ab Initio Methods for Electron Correlation in Molecules. In *Modern Methods and Algorithms of Quantum Chemistry*; Grotendorst, J., Ed.; NIC Series 1; John von Neumann Institute for Computing (NIC): 2000.
- [94] Schütz, M.; Hetzer, G.; Werner, H.-J. *J. Chem. Phys.* **1999**, *111*, 5691.
- [95] Schütz, M.; Werner, H.-J. *J. Chem. Phys.* **2001**, *114*, 661.
- [96] Schütz, M. *J. Chem. Phys.* **2002**, *116*, 8772.
- [97] Claeysen, F.; Harvey, J.; Manby, F.; Matta, R.; Mullholland, A.; Ranaghan, K.; Schütz, M.; Thiel, S.; Thiel, W.; Werner, H.-J. *Angew. Int. Ed.* **2006**, *45*, 6856.
- [98] Schütz, M.; Rauhut, G.; Werner, H.-J. *J. Phys. Chem. A* **1998**, *102*, 5997.
- [99] M. Schütz, unpublished implementation in MOLPRO.
- [100] Polly, R.; Werner, H.-J.; Manby, F. R.; Knowles, P. *Mol. Phys.* **2004**, *102*, 2311.
- [101] Weigend, F. *Phys. Chem. Chem. Phys.* **2002**, *4*, 4285.
- [102] Vahtras, O.; Almlöf, J.; Feyereisen, M. *Chem. Phys. Letters* **1993**, *213*, 514.
- [103] Casida, M. E. . In *Recent Advances in Computational Chemistry*, Vol. 1; Chong, D. P., Ed.; World Scientific: Singapore, 1995.
- [104] Runge, E.; Gross, E. *Phys. Rev. Lett.* **1984**, *52*, 997.
- [105] Christiansen, O.; Koch, H.; Jørgensen, P. *Chem. Phys. Lett.* **1995**, *243*, 409.

- [106] Kats, D.; Korona, T.; Schütz, M. *J. Chem. Phys.* **2006**, *125*, 104106.
- [107] Dreuw, A.; Head-Gordon, M. *Chem. Rev.* **2005**, *105*, 4009.
- [108] Görling, A. *Phys. Rev. Lett.* **1999**, *83*, 5459.
- [109] Furche, F.; Ahlrichs, R. *J. Chem. Phys.* **2002**, *117*, 7433.
- [110] Christiansen, O.; Jørgensen, P.; Hättig, C. *Int. J. Quantum Chem.* **1998**, *68*, 1.
- [111] Fliegl, H.; Köhn, A.; Hättig, C.; Ahlrichs, R. *J. Am. Chem. Soc.* **2003**, *125*, 9821.
- [112] Christiansen, O.; H.Koch,; Jørgensen, P.; Olsen, J. *Chem. Phys. Lett.* **1996**, *256*, 185.
- [113] Schreiber, M.; Silva-Junior, M. R.; Sauer, S.; Thiel, W. *J. Chem. Phys.* **2008**, *128*, 134110.
- [114] Sadeghian, K.; Schütz, M. *J. Am. Chem. Soc.* **2007**, *129*, 4068.
- [115] Knorr, A.; Daub, J. *Angew. Chem. Int. Ed. Eng.* **1997**, *36*, 2817.
- [116] Takeda, J.; Ohta, S.; Hirobe, M. *J. Am. Chem. Soc.* **1987**, *109*, 7677.
- [117] König, B.; Pelka, M.; Zieg, H.; Ritter, T.; Bouas-Laurent, H.; Bonneau, R.; Desvergne, J. P. *J. Am. Chem. Soc.* **1999**, *121*, 1681.
- [118] Hermann, D. T.; Schindler, A. C.; Polborn, K.; Gompfer, R.; Stark, S.; Parusel, A. B. J.; Grabner, G.; Köhler, G. *Chem. Eur. J.* **1999**, *5*, 3208.
- [119] Niemz, A.; Rotello, V. M. *Acc. Chem. Res.* **1999**, *32*, 44.
- [120] Trieflinger, C.; Rurack, K.; Daub, J. *Angew. Chem. Int. Ed. Eng.* **2005**, *44*, 2288.
- [121] Debreczeny, M. P.; Svec, W. A.; Marsh, E. M.; Wasielewski, M. R. *J. Am. Chem. Soc.* **1996**, *118*, 8174.
- [122] Debreczeny, M. P.; Svec, W. A.; Wasielewski, M. R. *Science* **1996**, *274*, 584.
- [123] Lukas, S.; Miller, E. E.; Wasielewski, M. R. *J. Phys. Chem. B* **2000**, *104*, 931.
- [124] Gong, X.; Ng, P. K.; Chan, W. K. *Adv. Mat.* **1998**, *10*, 1337.

- [125] Neiss, C.; Saalfrank, P. *Photchem. Photobiol.* **2003**, *1*, 77.
- [126] Neiss, C.; Saalfrank, P.; Parac, M.; Grimme, S. *J. Phys. Chem. A* **2003**, *107*, 140.
- [127] Bolboaca, M.; Iliescu, T.; Paizs, C.; Irimie, F. D.; Kiefer, W. *J. Phys. Chem. A* **2003**, *107*, 1811.
- [128] Pan, D.; Phillips, D. L. *J. Phys. Chem. A* **1999**, *103*, 4737.
- [129] Palafox, M. A.; Gil, M.; Nez, J. L.; Tardajos, G. *Int. J. Quantum Chem.* **2002**, *89*, 147.
- [130] Casida, M. E. . In *Recent Advances in Density Funktional Theory, Part I*; Chong, D., Ed.; World Scientific: Singapore, 1995.
- [131] Gross, E. U. K.; Dobson, J. F.; Petersilka, M. . In *Density Funktional Theory II*; Nalewajski, R., Ed.; Springer: Heidelberg, 1996.
- [132] Furche, F. *J. Chem. Phys.* **2001**, *114*, 5982.
- [133] Koch, H.; Jørgensen, P. *J. Chem. Phys.* **1990**, *93*, 3333.
- [134] Koch, H.; Jensen, H. J. A.; Jørgensen, P.; Helgaker, T. *J. Chem. Phys.* **1990**, *93*, 3345.
- [135] Christiansen, O.; Jørgensen, P.; Hättig, C. *Int. J. Quantum Chem.* **1998**, *68*, 1.
- [136] Michl, J.; Klessinger, M. . In *Excited states and Photochemistry of Organic Molecules*; VCH: New York, 1995.
- [137] Klessinger, M. *Angew. Chem. Int. Ed. Eng.* **1995**, *34*, 549.
- [138] Migani, A.; Olivucci, M. . In *Conical Intersections: Electronic Structure, Dynamics and Spectroscopy*; Domcke, W.; Yarkony, D.; Köppel, H., Eds.; World Scientific: Singapore, 2004.
- [139] Bernardi, F.; Olivucci, M.; Robb, M. A. *Chem. Soc. Rev.* **1996**, *25*, 321.
- [140] Boggio-Pasqua, M.; Bearpark, M.; Hunt, P. A.; Robb, M. A. *J. Am. Chem. Soc.* **2002**, *124*, 1456.
- [141] Bearpark, M.; Bernardi, F.; Clifford, S.; Olivucci, M.; Robb, M. A.; Smith, B. R.; Vreven, T. *J. Am. Chem. Soc.* **1996**, *118*, 169.

- [142] Sanchez-Galvez, A.; Hunt, P.; Robb, M. A.; Olivucci, M.; Vreven, T.; Schlegel, H. B. *J. Am. Chem. Soc.* **2000**, *122*, 2911.
- [143] Coto, P. B.; Sinicropi, A.; Vico, L. D.; Ferr, N.; Olivucci, M. *Mol. Phys.* **2006**, *104*, 655.
- [144] Ruiz, D. S.; Cembran, A.; Garavelli, M.; Olivucci, M.; Fuss, W. *Photchem. Photobiol.* **2002**, *76*, 622.
- [145] Zilbert, S.; Haas, Y. *J. Phys. Chem. A* **2002**, *106*, 1.
- [146] Abo-Riziq, A.; Grace, L.; Nir, E.; Kabelac, M.; Hobza, P.; de Vries, M. *Proc. Natl. Acad. Sci.* **2005**, *102*, 20 and references therein.
- [147] Schultz, T.; Samoylova, E.; Radloff, W.; Hertel, I. V.; Sobolewski, A. L.; Domcke, W. *Science* **2004**, *306*, 1765.
- [148] Sobolewski, A.; Domcke, W. *Chem. Phys. Lett.* **2003**, *294*, 73.
- [149] Sobolewski, A. L.; Domcke, W. *Phys. Chem. Chem. Phys.* **2004**, *6*, 2763.
- [150] Sobolewski, A.; Domcke, W.; Hättig, C. *Proc. Natl. Acad. Sci.* **2005**, *102*, 17903.
- [151] Perun, S.; Sobolewski, A. L.; Domcke, W. *J. Am. Chem. Soc.* **2005**, *127*, 6257.
- [152] Sobolewski, A. L.; Domcke, W. *Phys. Chem. Chem. Phys.* **2006**, *8*, 3410.
- [153] Gomez, I.; Reguero, M.; Boggio-Pasqua, M.; Robb, M. A. *J. Am. Chem. Soc.* **2005**, *127*, 7119.
- [154] Ahlrichs, R.; Bär, M.; Horn, H.; Kömel, C. *Chem. Phys. Lett.* **1989**, *162*, 165.
- [155] Schäfer, A.; Horn, H.; Ahlrichs, R. *J. Chem. Phys.* **1992**, *97*, 2571.
- [156] Eichkorn, K.; Weigend, F.; Treutler, O.; Ahlrichs, R. *Theor. Chem. Acc.* **1997**, *97*, 119.
- [157] Becke, A. D. *Phys. Rev. A* **1998**, *38*, 3098.
- [158] Becke, A. D. *J. Chem. Phys.* **98**, 1993, 5648.
- [159] Yamaguchi, Y.; Yokoyama, Y.; Mashiko, S. *J. Chem. Phys.* **2002**, *116*, 6541.

- [160] Tozer, D. J.; Amos, R. D.; Handy, N. C.; Roos, B. J.; Serrano-Andres, L. *Mol. Phys.* **1999**, *97*, 859.
- [161] Dreuw, A.; Fleming, G.; Head-Gordon, M. *J. Phys. Chem. B* **2003**, *107*, 6500.
- [162] Jamorski, C.; Foresman, J. B.; Thilgen, C.; Lüthi, H. P. *J. Chem. Phys.* **2002**, *116*, 8761.
- [163] Dreuw, A.; Head-Gordon, M. *J. Am. Chem. Soc.* **2004**, *126*, 4007.
- [164] Dreuw, A.; Weisman, J. L.; Head-Gordon, M. *J. Chem. Phys.* **2003**, *119*, 2943.
- [165] Tozer, D. J. *J. Chem. Phys.* **2003**, *119*, 12697.
- [166] Dreuw, A.; Fleming, G. R.; Head-Gordon, M. *Phys. Chem. Chem. Phys.* **2003**, *5*, 3247.
- [167] Rappoport, D.; Furche, F. *J. Am. Chem. Soc.* **2004**, *126*, 1277.
- [168] Hättig, C.; Weigend, F. *J. Chem. Phys.* **2000**, *113*, 5154.
- [169] Aquino, A. J.; Lischka, H.; Hättig, C. *J. Phys. Chem. A* **2005**, *109*, 3208.
- [170] Hättig, C.; Köhn, A. *J. Chem. Phys.* **2002**, *117*, 6939.
- [171] Köhn, A.; Hättig, C. *J. Am. Chem. Soc.* **2004**, *126*, 7399.
- [172] Kats, D.; Korona, T.; Schütz, M. manuscript in preparation.
- [173] Werner, H.-J.; Knowles, P. J.; Lindh, R.; Manby, F.; Schütz, M.; others., "MOLPRO, version 2006.2, a package of ab initio programs", 2006 see <http://www.molpro.net>.
- [174] Dunning, Jr., T. H.; Hay, P. J. . In *Methods of electronic structure theory*, Vol. 2; H. F. Schaefer III, Ed.; Plenum Press: , 1977.
- [175] Weigend, F.; Köhn, A.; Hättig, C. *J. Chem. Phys.* **2002**, *116*, 3175.
- [176] Barone, V.; Cossi, M.; Tomasi, J. *J. Chem. Phys.* **1997**, *107*, 3210.
- [177] Gerber, P. R. *J. Comput. Aided Mol. Des.* **1998**, *12*, 37.
- [178] Gerber, P. R.; Muller, K. *J. Comput. Aided Mol. Des.* **1995**, *9*(3), 251.

- [179] Laan, W.; van der Horst, M. A.; van Stokkum, I. H.; Hellingwerf, K. J. *Photchem. Photobiol.* **2003**, *78*, 290.
- [180] Dragnea, V.; Waagele, M.; Balascuta, S.; c. Bauer,; Dragnea, B. *Biochemistry* **2005**, *44*, 15978.
- [181] Masuda, S.; Hasegawa, K.; Ono, T.-A. *Biochemistry* **2005**, *44*, 1215.
- [182] Zirak, P.; Penzkofer, A.; Schiereis, T.; P.Hegemann,; Jung, A.; Schlichting, I. *Chem. Phys.* **2005**, *315*, 142.
- [183] Gauden, M.; Grinstead, J. S.; Laan, W.; van Stokkum, I. H. M.; Avila-Perez, M.; Toh, K.; Boelens, R.; Kapstein, R.; van Grondelle, R.; Hellingwerf, K. J.; Kennis, J. T. M. *Biochemistry* **2007**, *46*, 7405.
- [184] Zirak, P.; Penzkofer, A.; Hegemann, T. S. P.; Jung, A.; Schlichting, I. *Photchem. Photobiol. B.: Biol.* **2006**, *83*, 180.
- [185] Yuan, H.; Anderson, S.; Masuda, S.; Dragnea, V.; Moffat, K.; Bauer, C. *Biochemistry* **2006**, *45*, 12687.
- [186] Hasegawa, K.; Masuda, S.; Ono, T.-A. *Biochemistry* **2004**, *43*, 14979.
- [187] Masuda, S.; Hasegawa, K.; Ishii, A.; Ono, T.-A. *Biochemistry* **2004**, *43*, 5304.
- [188] Zirak, P.; Penzkfer, A.; Lehmpfuhl, C.; Mathes, T.; Hegemann, P. *Photchem. Photobiol. B.: Biol.* **2007**, *86*, 22.
- [189] Kita, A.; Okajima, K.; Morioto, Y.; Ikeuchi, M.; Miki, K. *J. Mol. Biol.* **2005**, *349*, 1.
- [190] Fukushima, Y.; Okajima, K.; Shibata, Y.; Ikeuchi, M.; Itoh, S. *Biochemistry* **2005**, *44*, 5149.
- [191] Okajima, K.; Fukushima, Y.; Suzuki, H.; Kita, A.; Ochiai, Y.; Katayama, M.; Shibata, Y.; Miki, K.; Noguchi, T.; Itoh, S.; Ikeuchi, M. *J. Mol. Biol.* **2006**, *363*, 10.
- [192] Laan, W.; Gauden, M.; Yeremenko, S.; van Grondelle, R.; Kennis, J. T. M.; Hellingwerf, K. J. *Biochemistry* **2006**, *45*, 51.
- [193] Unno, M.; Sano, R.; Masuda, S.; Ono, T.-A.; Yamauchi, S. *J. Phys. Chem. B.* **2005**, *109*, 12620.

- [194] Laan, W.; van der Horst, M.; van Stokkum, I.; Hellingwerf, K. *Photochem. Photobiol.* **2003**, *78*, 290.
- [195] Zirak, P.; Penzkofer, A.; Hegemann, P.; Mathes, T. *Chem. Phys.* **2007**, *335*, 15.
- [196] Kraft, B. J.; Masuda, S.; Kikuchi, J.; Dragnea, V.; Tollin, G.; Zaleski, J. M.; Bauer, C. E. *Biochemistry* **2003**, *42*, 6726.
- [197] Fukushima, Y.; Okajima, K.; Ikeuchi, M.; Itoh, S. *Photochem. Photobiol.* **2007**, *83*, 112.
- [198] Billeter, S. R.; Turner, A. J.; Thiel, W. *Phys. Chem. Chem. Phys.* **2002**, *2*, 2177.
- [199] Schütz, M.; Werner, H.-J.; Lindh, R.; Manby, F. R. *J. Chem. Phys.* **2004**, *121*, 737.
- [200] Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 1372.
- [201] A. Schäfer, C. H.; Ahlrichs, R. *J. Chem. Phys.* **1994**, *100*, 5829.
- [202] Kendall, R. A.; Dunning, T. H.; Harrison, R. J. *J. Chem. Phys.* **1992**, *96*, 6796.
- [203] Grinstead, J. S.; Hsu, S. T.; Laan, W.; Bonvin, J. J. M. A.; Hellingwerf, K. J.; Boelens, R.; Kapstein, R. *ChemBioChem* **2006**, *7*, 187.
- [204] Grinstead, J. S.; Avila-Perez, M.; Hellingwerf, K. L.; Boelens, R.; Kapstein, R. *J. Am. Chem. Soc.* **2006**, *128*, 15066.
- [205] Duvernay, F.; Chatron-Michaud, P.; Borget, F.; Birney, D.; Chiavassa, T. *Phys. Chem. Chem. Phys.* **2007**, *9*, 1099.
- [206] Schemmel, D.; Schütz, M. *J. Chem. Phys.* **2008**, *129*, 034301.
- [207] Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graph.* **1996**, *14*, 33-38.
- [208] Unno, M.; Masuda, S.; Ono, T.-A.; Yamauchi, S. *J. Am. Chem. Soc.* **2006**, *128*, 5638.
- [209] Toh, K.; van Stokkum, I.; Hendriks, K.; Alexandre, M. T.; Arnts, J. C.; Perez, M.; van Grondelle, R.; Hellingwerf, K.; Kennis, J. T. *Biophys. J* **2008**, *95*, 312.

- [210] Masuda, S.; Hasegawa, K.; Ono, T.-A. *Plant Cell Physiol.* **2005**, *46*, 1894.
- [211] Gauden, M.; Yeremenko, S.; Laan, W.; van Stokkum, I. H. M.; Ihalainen, J. A.; van Grondelle, R.; Hellingwerf, K. J.; Kennis, J. T. M. *Biochemistry* **2005**, *44*, 3653.
- [212] Hoffmann, M.; Wanko, M.; König, P. S. P.; Frauenheim, T.; Schulten, K.; Thiel, W.; Tajkhorshid, E.; Elstner, M. J. *Am. Chem. Soc.* **2006**, *128*, 10808.
- [213] Hazra, P.; Inoue, K.; Laan, W.; Hellingwerf, K. J.; Terazima, M. J. *Phys. Chem. B.* **2007**, *112*, 1494.
- [214] Kuhn, R.; Weygand, F. *Chem. Ber.* **1935**, *68B*, 1282.
- [215] Svoboda, J.; König, B.; Sadeghian, K.; Schütz, M. Z. *Naturforsch. B* **2008**, *63b*, 47.
- [216] Halkier, A.; Kloppe, W.; Helgaker, T.; Jørgensen, P.; Taylor, P. J. *Chem. Phys.* **1999**, *111*, 9157.
- [217] Werner, H.-J.; Manby, F. R.; Knowles, P. J. *J. Chem. Phys.* **2003**, *118*, 8149.
- [218] Pipek, J.; Mezey, P. G. *J. Chem. Phys.* **1989**, *90*, 4916.
- [219] Boughton, J. W.; Pulay, P. J. *Comput. Chem.* **1993**, *14*, 736.
- [220] Deglmann, P.; Furche, F.; Ahlrichs, R. *Chem. Phys. Letters* **2002**, *362*, 511.
- [221] Schäfer, A.; Huber, C.; Ahlrichs, R. *J. Chem. Phys.* **1995**, *102*, 346.

Erklärung

Ich erkläre hiermit an Eides statt, dass ich diese Arbeit selbst verfaßt und keine anderen als die angegebenen Hilfsmittel verwendet habe. Alle aus anderen Quellen direkt oder indirekt übernommenen Daten und Konzepte sind unter Angabe des Literaturzitats gekennzeichnet.

Regensburg, November 2008

Keyarash Sadeghian