Application and Development of Charge Transfer Simulations in Biological and Organic Systems

Zur Erlangung des akademischen Grades eines

DOKTORS DER NATURWISSENSCHAFTEN

(Dr. rer. nat.)

von der KIT-Fakultät für Chemie- und Biowissenschaften

des Karlsruher Institus für Technologie (KIT)

genehmigte

DISSERTATION

von

M. Sc. Daniel Holub

1. Referent: Prof. Dr. Marcus Elstner
2. Referent: Prof. Dr. Wolfgang Wenzel

Tag der mündlichen Prüfung: 06.05.2019

Eidesstattliche Erklärung

Die vorliegende Arbeit wurde unter Anleitung von Prof. Dr. Marcus Elstner in der Zeit von April 2016 bis März 2019 am Institut für Physikalische Chemie (IPC) des Karlsruher Instituts für Technologie (KIT) und, in der Zeit von Mai 2018 bis September 2018, am University College London (UCL) in der Gruppe von Prof. Dr. Jochen Blumberger durchgeführt.

Hiermit versichere ich, dass die vorliegende Arbeit selbstständig angefertigt wurde und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet wurden. Wörtlich oder inhaltlich übernommene Stellen wurden als solche kenntlich gemacht und die Satzung des Karlsruher Instituts für Technologie (KIT) zur Sicherung guter wissenschaftlicher Praxis in der gültigen Fassung wurde beachtet.

Karlsruhe, den 27.03.2019

Daniel Holub

Zusammenfassung

Diese Arbeit zeigt eine theoretische Untersuchung von lichtinduzierten Ladungstransferprozessen in der Proteinfamilie der Photolyasen/Cryptochrome, als prominentes Beispiel. Während Photolyasen lichtinduzierte DNA-Schäden reparieren, sind Cryptochrome Signa lmoleküle, welche die Lichtreaktion vermitteln. Beide haben eine homologe Proteinstruktur und einen Flavin-Adenin-Dinukleotid Kofaktor, der ihre biologischen Funktionen durch Ladungstransferprozesse ermöglicht. Diese weisen einen signifikanten nicht-gleichgewichts Charakter auf, welcher es für die etablierten Ladungstransfertheorien schwierig macht die richtige Kinetik vorherzusagen.
Daher wird in diese Arbeit eine hybride Multi-Skalen Simulationsmethode verwendet, die klassische und Quantendynamikansätze kombiniert und nicht auf der Annahme von Gleichgewichtsprozessen basiert ist. Dieses Multi-Skalen Methode wurde angewendet, gebenchmarkt und weiterentwickelt, um die Genauigkeit der Quantenbeschreibung zu untersuchen und zu erhöhen. Dies erhöhte auch die Übertragbarkeit auf Studien zum Ladungstransfer in anderen molekularen Systemen. Darüber hinaus bietet diese Methode eine praktikable Alternative zu teuren zeitaufgelösten Spektroskopie-Verfahren, hinsichtlich der Charakterisierung von Ladungstranferprozessen in den Photolyasen und Cryptochromen.

In experimentellen Studien mit *PhrB*, einer Photolyase von *A. tumefaciens*, war es nicht möglich, die Kinetik der Ladungstranferprozesse aufzulösen. Daher wurde in dieser Arbeit die Multi-Skalen Methode verwendet, um den Ladungstransfer in *PhrB* zu charakterisieren. Insbesondere das Vorhandensein von einem Tyrosin, anstelle von Tryptophan, im Ladungstransfer Pfad wurde untersucht. Daher wurden Molekulardynamik-Simulationen von Mutanten genutzt um eine Struktur-Funktionsbeziehung zu erstellen und ein getunnelter Ladungstransfer wurde vorgeschlagen, der die erste Aminosäure des Pfades involviert. Weitere Simulationen zeigten, dass das angrenzende Wasser einen Transfer entlang des energetisch ungünstigen Tyrosins ermöglicht, während ein Tryptophan den Rücktransfer erleichtert. *PhrA*, eine weitere Photolyase von *A. tumefaciens*, wurde untersucht, um die Ladungstranferprozesse entlang alternativer Pfade aufzuklären. Es wurden unterschiedliche Verhaltensweisen der Kinetik des Ladungstransfers beobachtet. Ein Pfad bietet einen schnellen Transfer, während der Andere eine bessere Stabilisierung der Ladung ermöglicht.
Der Ladungstransfer in Photolyasen und Cryptochromen löst weitere Prozesse aus, wie zum Beispiel einen Protonentransfer zum Kofaktor oder strukturelle Umlagerungen des Reaktionszentrums. In der *E. coli* Photolyase kann der Kofaktor protoniert werden, jedoch ist der Protonendonator und der Mechanismus unklar. Darüber hinaus zeigten experimentelle Mutationsstudien, dass die Einbringung eines Protonendonators in die Tasche des Kofaktors die Protonierung verhindert. Daher wurden klassische und Quantendynamik Simulationen durchgeführt und ein Protonentransfermech-

anismus bestimmt. Simulationen einer Mutante zeigten, dass die Flexibilität und die Coulomb-Wechselwirkungen der Tasche nicht durch den Austausch von einer oder zwei Aminosäuren verändert werden kann. Somit kann eine Protonentransferfähigkeit nicht eingeführt werden, auch wenn die Mutanten scheinbar die entscheidenden strukturellen Muster aufweisen.

Abstract

This thesis shows a computational study of light-induced charge transfer processes in the photolyase/cryptochrome protein family as a prominent example.
While photolyases repair photo-induced DNA damage, cryptochromes are signalling molecules, mediating the light response. Both have a homologous protein structure and a flavin-adenine-dinucleotide cofactor, which facilitates their biological function by means of charge transfer processes. These possess a significantly non-equilibrium character, which makes it difficult for the established charge transfer theories to predict the correct kinetics.
Therefore, this work employs hybrid multi-scale simulation methods that combine quantum and classical dynamics approaches, which do not underlay the assumption of equilibrium processes. This multi-scale method was applied in this thesis, benchmarked and further developed to observe and increase the accuracy of the quantum propagation. This increased the transferability toward studies of charge transfer in other molecular complexes. Furthermore, this framework is a viable alternative to expensive time-resolved spectroscopy methods, regarding characterisation of charge transfer processes in the photolyases and cryptochromes.

In experimental studies of *PhrB*, a photolyase from *A. tumefaciens*, it was not possible to resolve the kinetics of the charge transfer processes. Thus, the multi-scale method was used in this work to characterise the charge transfer in *PhrB*. Especially, the presence of a tyrosine instead of a tryptophan in the charge transfer pathway was investigated. Therefore, molecular dynamics simulation of mutants were used to derive a structure-function relationship and proposed a tunnelled charge transfer, involving the first amino acid of the pathway. Further simulations showed that the surrounding water enables a transfer along the energetically unfavourable tyrosine while tryptophan facilitates back transfers. *PhrA*, another photolyase from *A. tumefaciens*, was investigated to resolve the charge transfer processes along alternative pathways. Different behaviours of the charge transfer kinetics were observed. One pathway offers a fast transfer, while the other enables a better stabilisation of the charge.
The charge transfer in photolyases and cryptochromes triggers further processes such as a proton transfer to the cofactor or structural reorientations of the reaction centre. In the *E. coli.* photolyase the cofactor can be protonated, however, the proton donor and mechanism is unclear. Additionally, experimental mutation studies showed that insertion of a proton donor into the cofactor pocket hinders protonation. Thus, biased classical and quantum simulations were performed and a proton transfer mechanism was derived. Simulations of a mutant revealed that the flexibility and coulomb interactions of the pocket cannot be changed by replacing one or two amino acids. Thus, a proton transfer capability cannot be introduced even if the mutants seem to carry the crucial structural pattern.

Contents

Part I.

Introduction

1. Electron transfer in biological and organic systems

Chemical reactions involving an electron transfer (ET) happen permanently in every biological system and enable the live in the currently known form. Two essential required processes are the photosynthesis and respiration. Photosynthetic cells catalyse *via* the redox protein machinery the conversion of sunlight into reduction equivalents.[1–3] Transmembrane proteins construct an electron transport chain, which allows ET, proton transfer (PT) and coupled electron proton transfers, to enable the respiration in the membrane of mitochondria. Additionally, the transient docking of cofactors is required in the respiration, which will be oxidised and reduced by ET and PT reactions. For example, the complex I, in the respiration process offers one of the longest ET pathways around 300 Å, enabling the reduction of ubichinon to ensure the transport of four protons along the membrane.[4] These transfer reactions generate a transmembrane proton gradient, which allows the essential enzyme-catalysed reaction forming ATP out of ADP.[5–7] Several more remarkable ET processes have been discovered in the last decades. For example, microbes can grow filaments, used as an external electron transfer path along hundreds of Ångström, to offer an extracellular respiration type. Thus, biological ET transfer reactions are not restricted to intracellular processes.[8,9] Furthermore, ET along DNA is highly investigated and applications are derived, e.g. using DNA as nano wire.[10] Several biological processes have been clarified in which different enzymes use an ET along the DNA to detect lesions.[11] In many biological systems, a photo-induced ET enables their functionality. For example, photo-activated enzymes, such as Photolyases, can transfer an electron directly to damaged DNA to initiate the repair. In addition, the closely related Cryptochromes are also photo-activated and participate in various signal transtruductions. The function of both enzymes is based on ET reactions between a flavin-adenine-dinucleotide (FAD) cofactor and a pathway created out of neighbouring tryptophan amino acids.

These are just a few selected examples, which introduce that it becomes increasingly more and more important to understand biological ET to clarify those impressive processes and mechanisms. The experimental studies are challenging, which is for example based on the ultra-fast kinetics, in the pico- and milliseconds timescale. The expensive time-resolved spectroscopy methods, such as transient absorption spectroscopy, are essential to observe the processes. However, with these methods it is also difficult

to precisely identify the transiently formed intermediates. Additionally, the occurrence of side reactions or structural rearrangements can complicate the measurements. Theoretical calculations can be used to replace these measurements or to resolve those processes in combination with experimental observations. Therefore, theoretical methods shall be further improved to explain the observations (I), to solve unknown phenomena and to investigated mechanisms (II), and to support and expire new experiments, biological material and devices (III).

Based on a theoretical perspective view, the most challenging factor is the accurate description of the biological ET in the whole system for the requested timescale. Two different regimes are used to describe long-ranged ET in biomolecules: tunnelling and direct hopping mechanism. The tunnelled ET requires that the charge is initially localised at a defined donor and then transferred to a localised acceptor, happening roughly on a time scale of up to milliseconds. However, the distance between them does not allow a direct transfer and therefore an intermediate bridge is used as pathway.[12–15] Tunnelling occurs due to a large energy difference between the donor/acceptor and bridge moieties. Based on the energy difference, the bridge will not be occupied by the charge effectively and is just used as travelling structure. The bridge can be any part of the system e.g. backbone of the protein or non-covalent bond cofactors. The donor/acceptor can be metal atoms, small organic molecules, side chains of amino acids or DNA bases. The direct hopping ET mechanism, on the contrary, is possible for neighbouring molecules about 10 Å apart. Furthermore, these molecules need to be energetically similar, hence, thermal fluctuations can in- or decreases their energy to overcome the transfer barrier. This thermally activated energy increase allows the molecule to reach the transition state. Afterwards, the electron can be transferred to the neighbouring molecule which is determined as a hopping event. In the hopping mechanism, the charge is normally much faster transferred compared to tunnelling and the distance for the transfer are shorter, nevertheless, continuous hopps between molecules also enable a long-ranged ET.

Many theoretical approaches are developed to describe, as accurate as possible, the two different transfer mechanisms. Based on experimental well-studied ET processes, the theoretical approaches can be benchmarked and further improved. This enables to study unresolved processes or to support complex experimental studies with theoretical calculations. The established methods for theoretical calculations need to be improved and new methods need to be established. The already mentioned protein family of the Photolyases and Cryptochromes has been studied extensively on a theoretical and experimental level with coincident results. The construction of their ET pathway directly assumes that a hopping based description is able to predict the transfer successfully. However, well-established methods, such as the Marcus theory, which is based on the non-equilibrium character of the transfer, failed to reproduce the experimental observed ET transfer kinetics.

In the group of Elstner, a multi-scale method was developed, originally used to simulate charge transfers (CT) in DNA, and extended to enable simulations of the CT in proteins, reproducing the experimental measured transfer rates of Photolyases and Cryptochromes successfully.[16–18] The ET can also be simulated as an electron hole transfer in the opposite direction. The nature of the system facilitates the simulation of an electron or an electron hole transfer. However, both cases describe the same physical process and are determined as a CT, which will be used in following to simplify the description.
This multi-scale method allows a direct propagation of the charge by solving the time-dependent Schrödinger equation and a simultaneous simulation of the remaining system, using classical mechanics. This coupled dynamics directly enable that the changing position of the charge influences the surrounding environment and *vice versa*. The quantum nature of the CT, the complex structure and dynamics of the protein are considered simultaneously. The successful reproduction of the experimental results established this multi-scale method to study and characterise the CT of more members of the Photolyase and Cryptochrome protein family (PCF). Therefore, the established multi-scale method, can be used to calculate the CT reaction rates which allows the replacement of the expensive, experimental measurement. Additionally, it provides a molecular insight to clarify the relationship between the protein structure and the CT abilities of the protein.

1.1. Photolyase and Cryptochrome protein family – PCF

Photolyases (PL) and Cryptochromes (CRY) constitute a family of blue-light sensitive flavoproteins. The photolyase/cryptochrome family (PCF) is distributed over all three kingdoms of live ranging from bacteria to mammals.[19] Both proteins show a highly similar three-dimensional folded structure, which contains conserved structural motives, enabling their biological function: a flavine-adenine-dinucleotide (FAD) cofactor binding pocket and an apposition of several aromatic amino acids, which normally includes three tryptophan (Trp) side chains, called Triad.[20]
The PCF can be divided into three major groups: the CPD-Photolyase and the 6-4-Photolyase, which respectively repair cyclobutane pyrimidine dimers (CPD) and (6-4) pyrimidine pyrimidone adducts in UV-damaged DNA,[21] and the CRY, which has lost or just a reduced DNA repair activity but has gained a novel role in signalling.[19,22,23] PL are present in many more organisms than CRY and always used as a DNA repair system, while CRY are responsible for similar signalling pathways but contribute also to different signal cascades with respect to the organism. CPD-PL can be further separated into three subgroups, class I-III CPD-PL, showing variations in their structure.

The phylogenetic tree of the PCF, shown in Figure 1.1, gives a detailed overview of the relations between the different groups and several branches between CRY and PL.

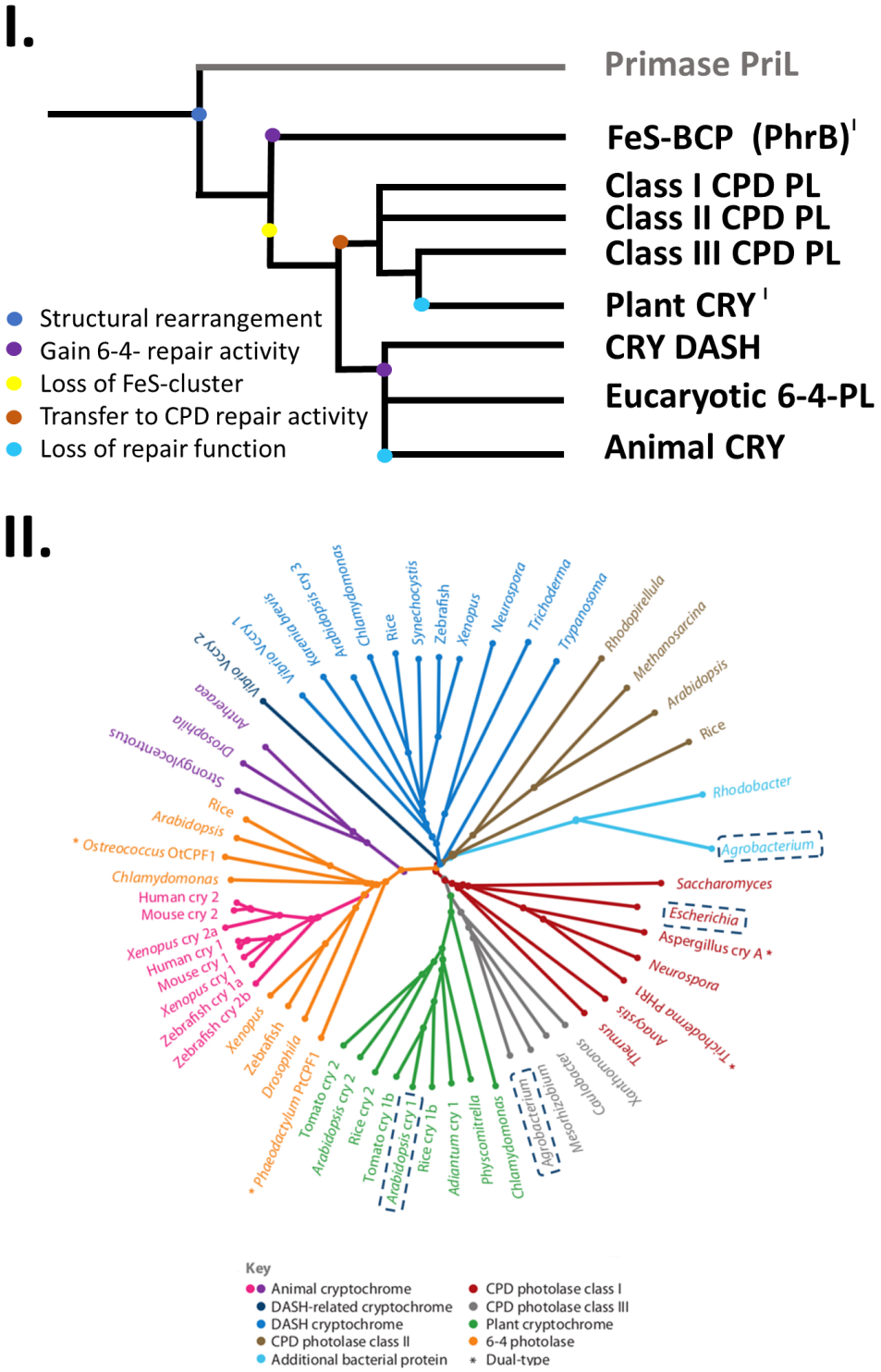

Figure 1.1.: Phylogenetic trees of the PL and CRY protein family: the upper picture I. shows the order of different subgroups and their direct relation. The coloured circles indicate important changes in their structural evolution, mentioned by the short descriptions. The four groups marked with a [I] were studied in this work. This Figure was generated in the same fashion as in Ref. 24–27 The lower picture II. shows the same subgroups and many organisms in which PL and CRY were investigated and identified. The surrounding dashed lines of four entries show the proteins, investigated in this work. The colour code, presented in the lower square, orders the colour to the subgroups. The Figure is taken out of Ref. 19 and modified.

Different hypotheses have been formulated concerning the evolution in the PCF.
Phylogenetic analysis of Tado et al. suggest an ancestral gene, encoding the class-I CPD-PL, for the CPF[28,29] which was further modified during the evolution of this protein family to form class-II and class-III CPD-PL. The hypothesis mentions that the class-I CPD-PL was also the pattern for 6-4-PL. Rearrangements of the structure *via* mutations might have changed the repair affinity and the mechanism. Thus, the 6-4-PL group can be seen as an evolutionary development derived from the initial class-I CPD-PL.[30]
The CRY can also be divided into more subgroups which all work as blue-light photoreceptors: animal CRY, plant CRY and CRY-DASH, which is structurally and functionally the closest relative to the PL, having slight DNA-repair ability of single strand-DNA.
On the contrary, another evolutionary guess, supported by phylogenetic studies of Lamparter et al.,[24,25] suggests a different relation between CRY and PL, based on an ancestor PL which is also related to the Primases. A photolyasen-homologues protein, called *PhrB*, was identified in the *Agrobacterium tumefaciens*, which is supposed to be closely related to the ancestor PL. A phylogenetic tree, shown in the part II. of Figure 1.1, displays the evolutionary development of the individual subgroups, assuming *PhrB* as pattern structure. The evolutionary hypothesis, derived by Lamparter, orders the evolution of the PCF by introducing a different ancestral class the iron-sulphur bacterial Cryptochromes and Photolyases (FeS-BCP). The eukaryotic PL, *PhrB*, and archaeal Primases (PriL)[31] show a related structure. PriL are enzymes synthesizing RNA oligonucleotides during the DNA replication. Both, PriL and FeS-BCP, contain an iron-sulphur (FeS) cluster, which is covalently liked to four cysteins (Cys) or rarely to histidine (His). Especially the region around the FeS-cluster shows the largest homologue fold (70% of FeS-BCP contain an FeS-cluster[27]). These similarities rises the guess that FeS-BCP are an intermediate group between the protein structure of PriL and PCF. FeS-cluster are often involved in ET processes[32] and it is known that the Fe-S cluster of human PriL is essential to initiate the primer synthesis, however, the complete mechanism is not resolved on an atomistic level.[33] The base excision repair requires two proteins, which also contain a FeS-cluster. It is supposed that these proteins, MutY a DNA glycosylase and EndoIII an Endonuclease, transfer an electron to the DNA to detect the DNA lesions.[11] Several further proteins are involved in DNA repair and contain a FeS-clusters. Thus, it may also have an impact in the repair by *PhrB* but it was replaced by other structural motives during the evolution of this protein family. On the contrary, the FeS-cluster could just influence the stability of the protein structure and be a remnant from the ancestor protein, because excluding it hindered the correct folding of the protein.[34] Beside the replacement of the FeS-cluster during the evolutionary development of the protein family, other also important structures might have been restructured.
In general, PL contain an antenna chromophore, which harvests light and transfers the light-activation to FAD. This establishes the active form of the protein, to fulfil its biological function. A rearrangement of the binding pocket of the antenna chromophore, beside

FAD, is supposed for *PhrB*.[24,25] A 6,7-dimethyl-8-ribityllumazine (DMRL) is present in *PhrB* and used as antenna chromophore. This chromophore is varied in different PL, for example an ethenylte-trahydrofolate (MTHF) in *E. coli* CPD-PL-I and a 8-hydroxy-5-deazariboflavin for CPD-PL-II. Another structural difference of *PhrB* is a C-terminal extension which is unusual for PL but usually present in CRY. The extension is required for the signal transduction in CRY but is not present in many other eukaryotic PL than *PhrB*.[35] Furthermore, it is proposed that FeS-BCP can also have a signalling functionality.[19,36]

1.1.1. Biological function of the PCF

Photolyases are enzymes which can bind and repair specific DNA damages. UV radiation at wavelengths of around 260 nm forms the DNA lesion by covalently linking of adjacent pyrimidines within the same strand which results in 70-80%[37] to a CPD four-membered ring and in 20-30%[38] to a 6-4 photo-adduct, which is illustrated in Figure 1.2. The cleavage of the different DNA lesions requires individual repair mechanism: the CPD can be repaired by one electron transfer (ET) between FAD and DNA, while the 6-4 photo-adduct requires an additional proton transfer (PT) reactions.[39]

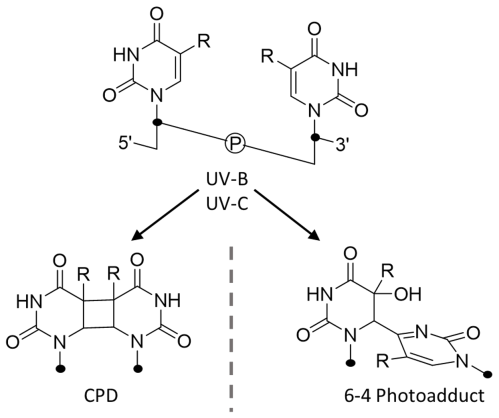

Figure 1.2.: Light-induced dimerisation of DNA bases: the top shows two DNA bases in a strand 5' to 3' the black circles indicate the sugar backbone of DNA. The second row shows the two possible products CPD and the 6-4 photo-adduct. *R* refers to H for uracil or CH_3 for thymine.

The binding pocket for the DNA in CPD-PL and 6-4-PL needs to offer a different environment. Thus, the individual repair mechanisms for the DNA lesion are ensured by the nature of the pocket.[40,41] Experimental studies showed that mutations in the DNA binding pocket of 6-4-PL can allow the repair of CPD but the other way around was not possible. Three mutations were necessary to convert the functionality of the 6-4-PL and even eleven mutations were tried to convert the CPD-PL into the 6-4-PL, but no repair could be observed.[42]

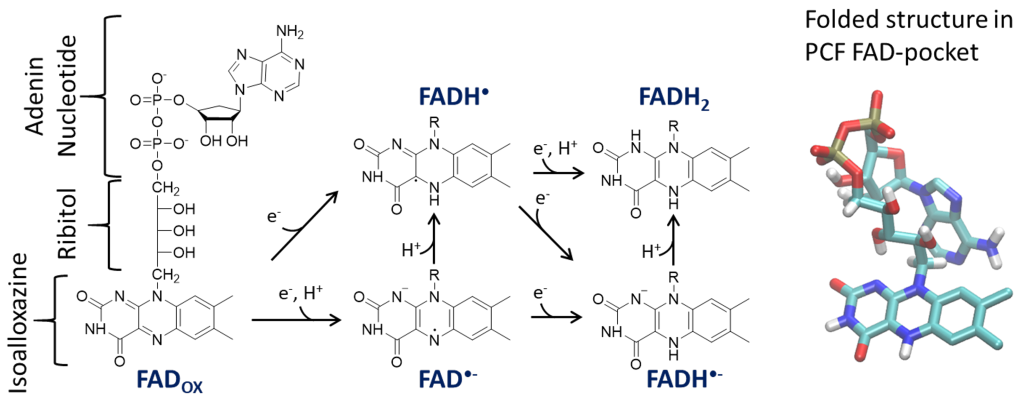

Figure 1.3.: Structural view of FAD: the left part of the picture shows the components building FAD: isoalloxazine ring system, ribitol sugar and adenine diculeotide. In the middle of the figure the possible PT and ET reactions and the resulting products, $FAD^{\bullet -}$ (anionic semiquinone), $FADH^{\bullet}$ (neutral semiquinone) and $FADH^{-}$ or the fully reduced $FADH_2$ (hydrochinone). The right part shows the typical U-stacked fold which is present in the PCF FAD binding pocket.

The PL FAD binding pocket, *in vivo*, shows a semiquinone, $FADH^{\bullet}$, or hydroquinone, $FADH^{-}$, which will become excited. Then, $FADH^{-*}$ is able to transfer an electron to the DNA, which will induce the repair of the DNA lesion. Therefore, the protonated FAD is requested for the biological repair function of this enzyme. The structural compounds of FAD and the details of the possible PT and ET reactions are presented in Figure 1.3. Experimental studies explored the kinetic for the ET forming $FADH^{-}$. Additionally, the ET kinetics involving a fully oxidised FAD, FAD_{ox}, and forming a $FAD^{\bullet -}$ were studied. The mentioned Triad is used as pathway for both ET reactions, forming the radical pair, $FADH^{-}$ or $FAD^{\bullet -}$ + $Trp^{\bullet +}$. Then, the positive charge is transferred along the Triad and the negative charge at FAD is stabilised by the nature of the pocket, which will be discussed later in more details. The timescale for a PT forming $FADH^{\bullet}$ out of $FAD^{\bullet -}$ was experimentally investigated and observed. Therefore, the pocket ensures that the essential oxidation state of FAD can be formed.[43]

The major role of CRY is to participate in the regulation of diverse biological responses as a signalling molecule. For instance, CRY contributes to the entrainment of the circadian rhythm[44], the space orientation by interaction with the magnetic field of the earth[19,20,45,46] which was especially shown for a CRY4 of birds[47] and the regulation growth or flowering state in plants.[48,49] It has been hypothesised that the signalling is based on the formation of the mentioned radical pair. Therefore, the semiquinone flavine, $FAD^{\bullet -}$ or $FADH^{\bullet}$ is required and needs to be stabilised. To enable this functionality, FAD also needs to become excited inducing the ET, such as in the PL, enabled by the same mechanism including an antenna chromophore.[50] The same conserved Triad allows the reduction of FAD^{*}_{ox} *via* ET reactions. Additionally, further PT reactions can form a transiently living $FADH^{\bullet}$ state in the presence of proton donors. Hence, CRY offers a FAD pocket with an environment supporting FAD to switch between redox states.

The exact signalling function and mechanism of CRY is not resolved completely. For example the CRYI phosphorylation, being partially self-mediated, is still unclear and an open question. Conformational changes might be required exposing the C-terminal end of the protein to ensure better interactions with other proteins such as COP1, which is a well-characterised repressor of photomorphogenesis, shown by Bouly[51]. The localised negative charge at FAD after the ET or at the enclosed proton donor, when further PT is considered, can probably induce structural reorientations of close α-helices.[52] Thus, it can be supposed that rearrangements in the protein structure of CRY are possible and essential for the function, compared to the PL.[53] Figure 1.4 summarises the observed PT and ET processes in PL and CRY and the corresponding functions to present the differences and similarities.

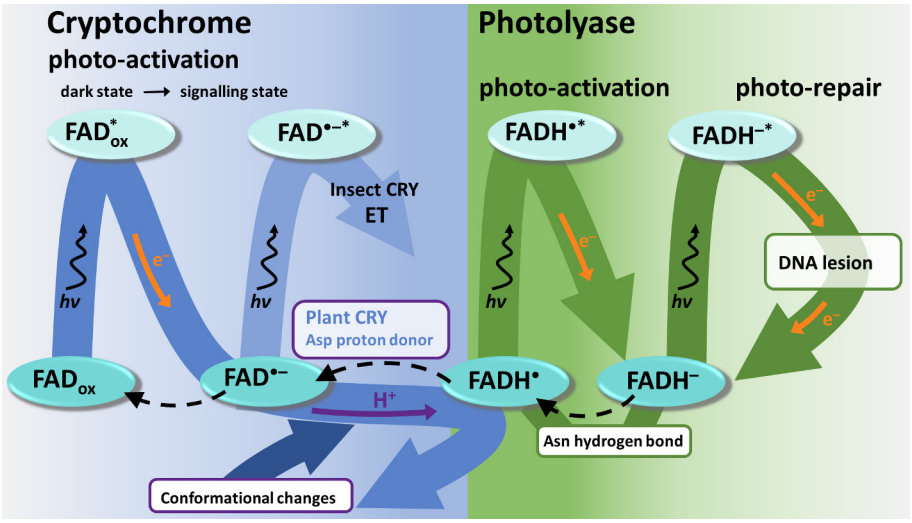

Figure 1.4.: Schematic view of ET and PT processes in PCF: the left side, coloured in blue, shows the photo-induced ET in CRY which shows *in vivo* fully oxidised FAD_{ox} and undergoes at lest one reduction forming an radical pair to ensure its signalling function. The right side, coloured in green, shows the similar photo-induced ET forming a $FADH^-$ which can become also excited to transfer the electron to the DNA lesion. The figure is inspired by Ref. 19, however, modified and extended.

1.1.2. FAD binding pocket

A similar fold of the FAD pocket is conserved in most PCF proteins to a large extent.[54] However, despite the large similarity, the pocket is obviously supposed to offer different interaction patterns which enables the mentioned, varying functions: in most CRY, a $FAD^{\bullet-}$ radical is formed while in PL a $FADH^-$ needs to be stabilised which needs to be ensured by the environment of the FAD pocket. The ET pathway exists in both proteins but the PT ability and the finally stabilised oxidation state, offered by the pocket, are not the same. The structure of the pocket highlights one specific amino acid in a position close to the N_5 of FAD, which differs between members of the PCF: in most plant CRY, e.g. CRYI from Arabidopsis thaliana (CRYI), an aspartate (Asp) is found in this

position,[55,56] but a non-titratable residue is often present, such as a cysteine (Cys) in entomic CRY.[57] In CRY-DASH of a cyanobacterium,[58] which shows a high similarity in structure and function to the PL, an asparagine (Asn) is found at this position. An Asn in this position is also present in PL, e.g. in class-I CPD-PL of *E. coli*[59] and *A. nidulans*.[60] The impact of this amino acid has been studied by experimental site-directed mutagenesis, showing that it affects the behaviour of the proteins decisively: in CRYI, the Asp acts as proton donor to form $FADH^{\bullet}$ from a photo-reduced semiquinone $FAD^{\bullet -}$. A mutation of Asp into Cys, modelling the FAD pocket of an entomic CRY, prevents the FAD protonation.[61] This shows that in CRYI this protonation is ensured by Asp396 and the presence of Cys prevents the formation of $FADH^{\bullet}$. Surprisingly, the mutation of Asp396 into Asn leads to the formation of $FADH^{\bullet}$ despite the lack of an obvious proton donor, which indicates the presence of an alternative PT pathway. Additionally, the Asp to Asn mutant in CRYI showed a PL-like DNA repair activity, which in fact requires the stabilisation of $FADH^{-}$.[55]

Conversely, the presence of Cys in entomic CRY, such as in *Drosophila melanogaster* CRY, inhibits the protonation of $FAD^{\bullet -}$.[62] In entomic CRY, the mutation of the Cys into Asp, albeit offering a suitable proton donor to FAD, does also not lead to the formation of $FADH^{\bullet}$.[63] On the contrary, a mutation of Cys into Asn shows a stabilisation of protonated $FADH^{\bullet}$, which does not affect the photo-receptor function.[63]

In the PL of *E. coli* and *A. nidulans*, as observed in the entomic CRY, the mutation of Asn into Asp highlights, that the inserted proton donor does not enable a PT to form $FADH^{\bullet}$.[43,64] After photo-reduction, the *E. coli* PL Asn to Asp mutant, called PL-N378D, shows a slightly different UV-VIS absorption spectrum compared to the spectrum for $FAD^{\bullet -}$. The presence of a different FAD state, called FAD^{x}, is observed and especially the absence of $FADH^{\bullet}$. It was hypothesised that this state indicates a strong hydrogen bond network between Asp and FAD N_5 or a shortly living $FADH^{\bullet}$. Surprisingly, the presence of Asp prevents the stabilisation of $FADH^{\bullet}$, thus, Asp cannot act as proton donor in the pocket of PL-N378D and inhibits any other PT pathway.

Additional, experimental studies investigated the impact of the Asn in PL. The replacement of Asn into Ser in the *E. coli* PL showed a drop of the *in vitro* DNA repair activity from 50% for wild type to less than 1% for the mutant.[65] This underlines the dominant role of Asn in DNA repair. Furthermore, combined spectroscopic and computational studies on *E. coli* PL showed the strong hydrogen bonding between $FADH^{-}$ and Asn, stabilising the $FADH^{-}$ state.[64,66]

Taken together, these results highlight that inserting an Asn by mutation allows the formation of $FADH^{\bullet}$ in the examined PCF. However, inserting an Asp instead of a Cys or Asn hinders the stabilisation of $FADH^{\bullet}$, which is also summarised in Table 1.1. Asn increases the stabilisation of $FADH^{\bullet}$ and could be essential for the DNA repair. The mutated CRYI, with the inserted Asn, offered a repair function.

The reported observations highlight the fine balance of the individual amino acids, sta-

bilising the active form of the protein.[67] It appears that the nature of the residue, interacting with the FAD N_5 which may be Asp, Asn or Cys, is strongly related to the biological most relevant FAD protonation state. A mutation does not necessarily result in a functional conversion of the mutants. Consequently, other structural or dynamical parameters, such as the formation of a PT path, influence the protonation and redox state of FAD in the PCF, because the insertion of an proton donor did not enable PT but the insertion of an Asn stabilising the product allowed PT.

Table 1.1.: Summarised effects of mutations in FAD pocket: the mentioned experimental observed effects of mutations involving the amino acid enclosed to FAD N_5.

Protein	Amino acid enlcosed to FAD N_5	stabilised FAD state after photo-induced ET
entomic CRY	Cys	$FAD^{\bullet -}$
	Cys $\rightarrow$ Asp	$FAD^{\bullet -}$
	Cys $\rightarrow$ Asn	$FADH^{\bullet}$
plant CRYI (*Arabidopsis thaliana*)	Asp	$FADH^{\bullet}$
	Asp $\rightarrow$ Cys	$FAD^{\bullet -}$
	Asp $\rightarrow$ Asn	$FADH^{\bullet}$
class-I CPD-PL (*E. coli*)	Asn	$FADH^{\bullet}$
	Asn$\rightarrow$ Asp	FAD^{x}

1.1.3. Electron transfer pathway

Three Trp residues at position 306, 382 and 359 of *E. coli* class-I CPD-PL, (PDB:1DNP) constitute the first identified ET cascade of the PCF, named Triad.[59] The protein structure allows the Trp to assume a position in which the aromatic side chains are in a π-π-stacked position. This ensures that the π-orbitals between the Trp can overlap. The structural orientation of aromatic amino acids is conserved in nearly all members of class-I-III CPD-PL, CRY-DASH proteins, plant CRYs, animal CRYs and eukaryotic 6-4 PL. Figure 1.5 illustrates the construction of the Triad and the orientation of the individual Trp residues and FAD. The overlap between the orbitals and their distance correlate with the strength of their electronic coupling, which determines the ET probability between those aromatic molecules.

Analysis of the crystal structures in combination with site directed mutagenesis identified that the amino acids, forming the Triad, constitute to the ET cascade and ensures the photo-activation of the proteins. In mutagenesis studies a relevant Trp or Tyr was replaced by phenylalanine (Phe) or alanine (Ala), which results in a slower or even blocked photo-reduction. However, *in vivo* the functionality of the PCF member is not completely inhibited by the mutation in the Triad, for example a CRY was still able to work as signalling molecule.[68] Thus, alternative mechanism may exist and are still under debate and the interpretation of mutant results can be hampered by the possibility of parallel pathways.[69,70] Based on these assumptions the clarification of the ET in these proteins is still highly interesting.

The whole ET kinetics along the Triad are time-resolved for some PCF and two prominent examples are the *E. coli* PL and *Arabidopsis thaliana* CRYI, investigated of Zhong et. al[39,71,72] and Kottke et al.,[61,73,74] respectively. The photo-activation of both proteins can be separated into the following stepwise ET processes, schematically shown in Figure 1.5: the excitation of FAD and the first ET between FAD and Trp happens in less than 1 ps, followed by two ET along the next Trp of the Triad, happening in around 200-300 ps. The transfer times are always in that regime but differ between individual proteins slightly. The ET creates the separated charge state and the already mentioned radical pair, $FAD^{\bullet -}$+ $Trp^{\bullet +}$. The ending of the Triad exposes the last $Trp^{\bullet +}$ to the solvent, which allows further stabilisations derived by the environment. A deprotonation of the last Trp was also observed which will increase the lifetime of the radical pair until the back transfer of the charge happens. This is also experimentally resolved, for example in *E. coli* PL the protonated or deprotonated happens in the nano- or millisecond regime, respectively.[43]

The aromatic side chain of Trp is chemically ideally suited for ET processes based on its low redox potential (ca. 0.6 V at pH 7)[75] and a stable radical state for the deprotonated form ($Trp^{\bullet +}$). In class-I and class-III CPD-PL, the involvement of additional Trp residues in photo-reduction has been shown. The class-II CPD-PLs have another Trp

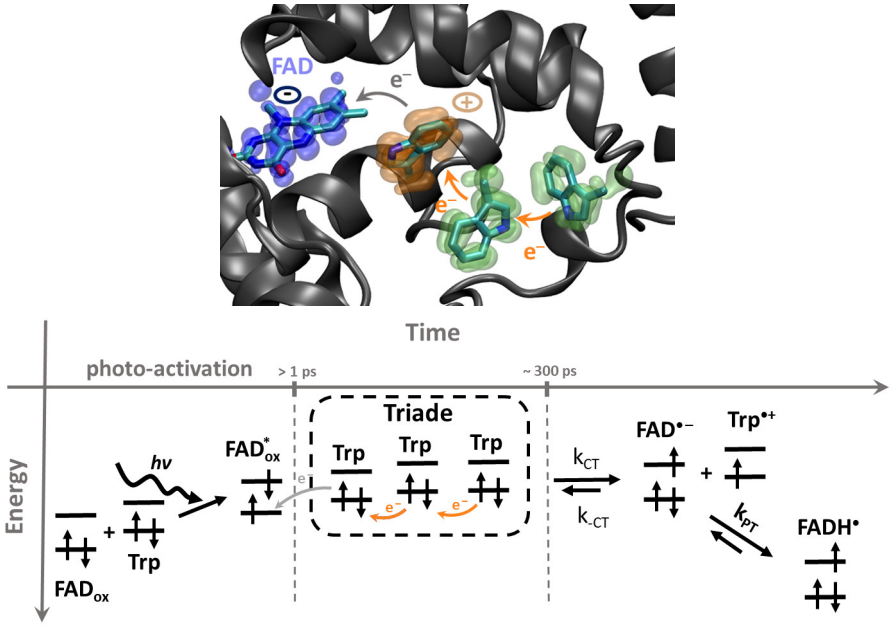

Figure 1.5.: ET between FAD and Triad in PCF: the upper part shows FAD, Trp residues and their orientations at a molecular view. The orbital of FAD in which the electron will remain after the transfer is presented and coloured in blue, the orbital of the first Trp in orange, in which an electron hole remains after the transfer and the neutral orbitals of the other two Trp residues in green. A schematic view of the transfer is presented in the lower part of the figure, which orders roughly the energy levels of the involved orbitals of FAD and Trp. The three individual processes are shown: I. shows the excitation of FAD and the ET between Trp and the single occupied orbital of FAD^*_{ox}, happening in less than 1 ps (photo-activation), II. shows the propagation along the Triad to the third position and III. shows the stabilised radical pair $FAD^{\bullet-}$ and $Trp^{\bullet+}$.

Triad, which is conserved among this group[76] or in the class-III CPD-PL, for example *PhrA* from *A. tumefaciens*, the first Trp, Trp384 of the classical Triad, is linked to a second Triad. Furthermore, in plant CRY or amphibian CRY additional Trp residues were identified, which can probably extend the Triad.[71,77–79] For example in plant CRYI of *Arabidopsis thaliana* additional Trp (Trp334, Trp379) enclosed to the Triad may enable to transfer the charge to a different part of the protein.[70]

Additionally, in several CPD class-I and class-II PL, Tyr residues are also involved in ET. For example, in the class-I CPD-PL from *Anacystis nidulans* a Tyr radical is formed within 50 μs after excitation of FAD, detected by ultrafast spectroscopy.[80] In *Methanosarcina mazei* class-II CPD-PL, a Tyr residue is even required for full photoreduction, which is also the case in the *Xenopus laevis* 6-4-PL showed by electron paramagnetic resonance.[40] In CRYI of *Arabidopsis thaliana*, a non-covalently bond adenine-triphosphate (ATP) was supposed to form a pathway to a Tyr402, which would be supported by the experimental measured Tyr radical formation without directly known pathway.[53,73,78] Tyr is most of the time placed at the ending of the Triad, because positively charged Tyr requires deprotonation to be stabilised. Therefore, ET involving a Tyr is normally present at the surface of the protein or enclosed to a proton acceptors,[77,81] whereas, ET involving Trp can occur in the centre of a protein or at the periphery. Thus,

the protein architecture of PCF uses Tyr and Trp in selected positions to ensure the requested ET abilities of the proteins. The strong conversation of the Triad or alternative extensions of the pathway assumes that the presence of Trp or Tyr residues is not a random process in PCF.

However, the mentioned FeS-BCP show a conserved Tyr,[25] instead of a Trp as a direct electron donor, in the normally conserved Trp Triad. Two Trp complete this pathway to reach the protein surface. Mutational studies from Lamparter et al.[34] were based on the photo-activation in *PhrB*, a FeS-BCP PL, and showed that the last two Trp (Trp324 and Trp390) are essential for the photo-reduction. The replacement into Phe of each neglects the formation of $FAD^{\bullet-}$, however, the replacement of the first member of this alternative Triad, Tyr391, into Phe does just slightly influence the formation of $FAD^{\bullet-}$.[26,34] Therefore, *PhrB* rises questions about the ET properties of the PCF. The presence of a Tyr and not prevented ET due to an inserted Phe points out that the environment may play a certain role in this PL. This reiterates the assumption that the general replacement of Trp into Phe hinders ET in all PCF, and therefore, a tunnelled ET can be considered. The surrounding environment of the pocket needs to be taken into account, which can also influence the ET highly. The impact of the solvent was already investigated, which revealed that the fast transfer rates of the positive charge of $Trp^{\bullet+}$ along the Triad is driven by the solvent.[18]

2. Motivation

The aim of this work was to solve experimental raised questions, which are based on electron transfers in biological systems, using theoretical approaches. Therefore, an established multi-scale simulation method shall be used to analyse the electron transfer in two Photolyases form *A. tumefaciens* (*PhrB* and *PhrA*) to replace the experimental expensive and challenging time-resolved spectroscopy measurements.
The simulations shall reveal an atomistic view to resolve a structure-function relationship, which can explain the observations of the experimental mutation studies. Furthermore, the electron transfer pathway in *PhrB*, which differs compared to other Photolyases, and alternative branched pathways in *PhrA* shall be investigated.

The electron transfer can induce a proton transfer and structural rearrangement in several members of the Photolyase/Cryptochrome protein family. Therefore, simulations shall explore the relation between the protein structure and the proton transfer capability to derive a mechanism for the protonation. The results of the simulations shall be used to clarify the understanding of the connection between the similar protein architecture of Photolyases and Cryptochromes and the corresponding different functionality.

Furthermore, the applied multi-scale method shall be improved. Therefore, a test system shall be designed to benchmark the quantum propagation method. Additionally, a different propagation scheme shall be implemented and benchmarked to increase the transferability, which allows the investigation and characterisation of the electron transfers in further unknown systems.

Part II.

Theoretical Background

3. Quantum chemistry

Computational chemistry contains a broad field of various useful methods, which all have their strengths and disadvantages. It is possible to solve a wide range of open questions, based on differing topics, with completely different methods. One of the most important factors is the size of the system, due to the fact that in some methods the calculation time scales with up to the 3rd order, or even higher, of the system size.

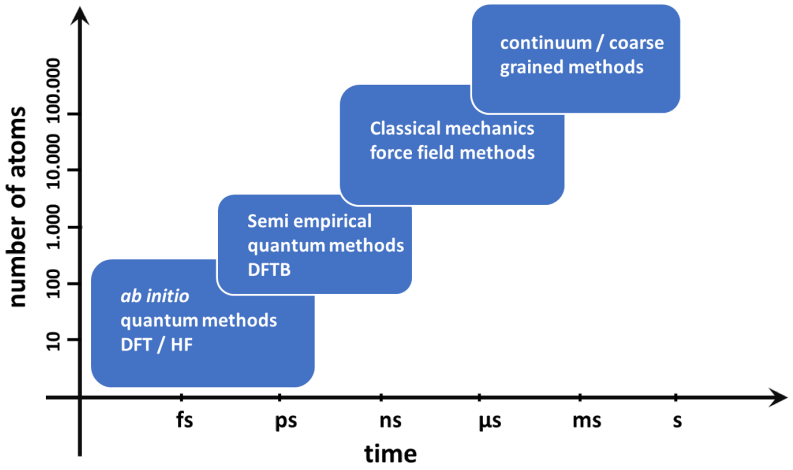

Figure 3.1.: Computational methods in relation with the system size and simulation time: the x-axis shows the computable simulation time and the y-axis the size of the system. The blue squares represent the methods which are normally applied to solve the requested open issue in an acceptable calculation time. The proposed methods are written on the blue squares.

Afterwards, the question of the time regime is important: is a simulation of dynamical structures required or is a calculation based on single geometries sufficient? This needs to be determined before a successful solution can be achieved. Therefore, the number of calculations in- or decreases, which directly affects the size of the system, which can be increased or needs to be reduced.
Furthermore, questions of surrounding structures beside the focused molecules need to be taken into account and their impact needs to be evaluated. For example the representation of the environment/solvent which can be modelled by a continuous model, an *in vacuo* calculation or an explicit representation of the molecules. On one hand the more explicit calculation of the environment increases the system size and therefore the calculation time. On the other hand, neglecting important aspects, possibly induced by the environment, can cause wrong results.

Thus, Figure 3.1 visualises the relation between the system size, the time regime of the simulation and proposes a possible computational method to explore the open question of interest. Some of the mentioned methods will be shortly presented in the next sections, the general assumptions will be explained and the strengths and failures addressed. Calculating the system with so called *ab inito* quantum methods (QM) gives full insight into all chemical properties, however, it limits simulation time and system size strongly. The basic assumptions of one method, Density Functional Theory, are presented in the following section 3.2.
Afterwards, a cheaper semi-empirical approach will be introduced in section 3.3. This can be used to decrease the calculation time and therefore increase the system size. The basic idea is that *ab inito* QM methods are used once for the parameterisation of the method and then a similar accuracy can be obtained.
On the contrary, neglecting the QM calculation and using just classical mechanics to calculate the interactions, called molecular mechanic (MM) methods (see section 4.1), allows to deal with much larger systems with up to millions of atoms. However, no quantum chemical processes are taken into account and the expression of atoms is further approximated to simple models. Additionally, combinations of QM and MM methods are established subdividing the system to gain a selected part described by a QM or semi-empiric method, called QM/MM simulation.

All quantum mechanical information about stationary properties of interest, such as the total energy, E_0, or ionisation potential (IP) of a defined molecular system is contained in the solution of the time-independent Schrödinger Equation (SE). Ψ is the many-body wave function for the system, which consists of n electrons with their defined coordinates r_{el} and momenta p_{el} and of N nuclei and their coordinates R_N and momenta P_N.

$$\hat{H}_{\mathrm{mol}}\Psi(r_n, R_N) = E_{\mathrm{mol}}\Psi(r_n, R_N) \tag{3.1}$$

$\hat{H}$ is the time-independent Hamiltonian and E_{mol} are the corresponding eigenvalues, defining the total energy of the system. $\hat{H}$ contains the following terms, describing the interactions of the included nuclei and electrons:

$$\begin{aligned}
T_{\mathrm{el}} &= \sum_{i=1}^{n} \frac{p_{\mathrm{el}}^2}{2m_{\mathrm{el}}} \\
T_{\mathrm{nuc}} &= \sum_{k=1}^{N} \frac{P_N^2}{2M_{\mathrm{N}}} \\
V_{\text{el-el}} &= \sum_{i}^{n} \sum_{j>i}^{n} \frac{e^2}{|r_i - r_j|} \\
V_{\text{el-nuc}} &= \sum_{i}^{n} \sum_{k}^{N} \frac{z_k e^2}{|r_i - R_N|} \\
V_{\text{nuc-nuc}} &= \sum_{k}^{N} \sum_{l>k}^{N} \frac{z_k z_l e^2}{|R_k - R_l|}
\end{aligned} \tag{3.2}$$

The mass of the nuclei, M_{N}, or electron, m_{el}, and their momenta enters the calculation of the kinetic energies, T, and due to their charges coulomb terms, V, appear, describing repulsive or attractive interactions.

3.1. Born-Oppenheimer approximation

The Born-Oppenheimer (BO) approximation allows a simplification which is based on the large difference in the masses of electrons and nuclei ($m_{\mathrm{n}}/M_{\mathrm{N}} < 10^{-3}$). Thus, the electrons move much faster and therefore follow the positions of the nuclei instantaneously. Hence, the movements of electrons and nuclei can be seen as decoupled and the resulting electronic Hamiltonian, $\hat{H}_{\mathrm{el}}$, depends on the nuclei positions parametrically and contains the following terms:

$$\hat{H}_{\mathrm{el}} = T_{\mathrm{el}} + V_{\text{el-el}} + V_{\text{el-nuc}} \tag{3.3}$$

The solution of the time-independent electronic SE describes the state of the electrons in an electrostatic field, induced by the position of the nuclei. Nevertheless, the solution of the SE in the BO still has the many-body interactions between the electrons neglecting a possible analytic and exact solution for systems having more than one electron, which is just possible for one-electron systems such as H_2^+.

Many quantum mechanical methods try to find the best possible approximation for Ψ to obtain the total energy of the system, using the variational principle. This is based on the assumption that any used trial wave function Ψ_{trial} will yield a corresponding energy, E_{trail}, which is higher than the exact ground state energy E_0:

$$\langle \Psi_{\mathrm{trial}} | \hat{H} | \Psi_{\mathrm{trial}} \rangle = E_{\mathrm{trial}} \geq E_0 = \langle \Psi_0 | \hat{H} | \Psi_0 \rangle \tag{3.4}$$

Generally, there are two different classes of quantum mechanical methods based on this assumption to solve the SE: the Hartree-Fock (HF) theory and the Density Functional Theory (DFT). In the next part, DFT will be explained further, because DFT and additionally an approximative semi-empiric DFT-based method is used in this work.

3.2. Density Functional Theory – DFT

Density Functional Theory (DFT) is based on the proof of Hohenberg and Kohn, who showed a direct one-to-one relation between the electron density $\rho(r)$, the external potential and the electronic wave function $\Psi(r,R)$. Therefore, the wave function, Ψ, can be expressed by a functional of only the density, $\Psi(r;R) \rightarrow \Psi[(\rho)]$. Thus, the quantities of interest, such as the total energy of the system, can be calculated as a functional of

the density, $\langle|\hat{A}|\rangle \rightarrow A[\rho(r)]$.[82]

This results in a system depending on just three spatial coordinates, which define the electron density, in contrast to the wave function, depending on 3n coordinates. Furthermore, systems with more electrons do not increase the complexity by introducing more variables, because the electron density is independent of the system size and therefore defined by the square of Ψ integrated over $N-1$ electron coordinates.[83] However, the exact functional, connecting the electronic density $\rho_0(r)$ to the corresponding energy E_0, is required but still unknown.

Many approaches tried to overcome this problem and some are described in the following. Generally, the energy can be decomposed into three parts:

$$E[\rho] = T[\rho] + E_{\text{el-nuc}}[\rho] + E_{\text{el-el}}[\rho] \tag{3.5}$$

$T[\rho]$ is the kinetic energy functional of the electrons, $E_{\text{el-nuc}}[\rho]$ the interaction between the external field, induced by the position of the N nuclei and the n electrons, and $E_{\text{el-el}}[\rho]$ is the interaction between electrons. Therefore, $E_{\text{el-nuc}}[\rho]$ is the only exactly known functional, which only dependents on the position of the nuclei R_a and their partial charge, Z_a:

$$E_{\text{el-nuc}}[\rho] = -\sum_a \int \frac{Z_a \rho(r)}{|R_a - r|} dr \tag{3.6}$$

The first two parts of Eq. 3.5 are tried to be modelled by several density dependent functionals. The first attempt was done by Thomas[84] and Fermi[85] who used the electron density and not the wave function to yield information about their system, called Tomas-Fermi model. Here, $E_{\text{el-nuct}}[\rho]$ and $E_{\text{el-el}}[\rho]$ are calculated only on a classical level and the kinetic energy is based on the uniform electron gas, a model system having a constant electron density. This results in an approximative kinetic energy and fails to reproduce energies for inhomogeneous electron distributions. Therefore, this model does not yield right energies for chemical bond formation, thus, it neglects the presence of molecules due to its rough estimation of the kinetic energy. Further improvements also failed to reproduce the accuracy of wave function methods. The inclusion of additional terms, depending on the derivative of ρ, did not solve the problem. This extension can be interpreted as assuming a heterogeneously distributed electron gas and performing a Tylor-like expansion with the density as variable.[83]
However, this shows that in general it is possible to derive the energy by just using ρ and for the homogeneous electron gas it can be exactly described for a high- or low-density limit.[86]

To overcome this issue, Kohn and Sham applied an auxiliary system of non-interacting electrons[87] and introduced molecular orbitals (MO) to DFT. MO are used in HF theory normally and orthonormal by construction ($\langle\phi_i|\phi_j\rangle = \delta_{ij}$), therefore, Eq. 3.5 be-

comes Eq. 3.7 in which the kinetic energy for non-interacting electrons is represented as $T_S[\phi[\rho]]$, which depends on the MO, ϕ_i.

$$
\begin{aligned}
E[\rho] &= T_S[\phi[\rho]] + E_{\text{el-nuct}}[\rho] + J[\rho] + E_{xc}[\rho] \\
&\text{with} \\
T_S[\phi[\rho]] &= \sum_i^{occ} \langle \phi_i | -\frac{1}{2}\nabla^2 | \phi_i \rangle \\
J[\rho] &= \frac{1}{2}\int \frac{\rho(r)\rho(r')}{|r-r'|} dr dr' \ .
\end{aligned}
\tag{3.7}
$$

Here, $E_{\text{el-el}}$ is split into $J[\rho]$, the classical coulomb energy of a charge density, which can be calculated using the introduced, non-interacting auxiliary system and other parts, which can not be calculated using the auxiliary system, being stored in E_{xc}, the exchange-correlation functional. In the same fashion, the kinetic energy of the system is separated in to two parts: the one which is calculated *via* the auxiliary system and the missing one taking the electron-electron correlation into account.
Due to this, the last term, E_{xc}, contains the difference between the used non-interacting auxiliary system and the real system, shown in Eq. 3.8.

$$
E_{xc}[\rho] = (T[\phi[\rho]] - T_S[\phi[\rho]]) + (E_{\text{el-el}}[\rho] - J[\rho]) \tag{3.8}
$$

Many approximations were made to reproduce the exact exchange-correlation functional, $E_{xc}[\rho]$, which yield successful results, shortly described in the following: the local-density-approximation (LDA) takes only the value of ρ in each point into account. Hence, it is based on the assumption of the homogeneous electron gas and more useful for solids than for molecules.
Further, the derivative of the density is included in generalised gradient approximations (GGA), which showed a large increase of the accuracy of the results. There is no direct way to improve $E_{xc}[\rho]$ but known conditions of the exact $E_{xc}[\rho]$ are reproduced in an approximative form. LDA and GGA cannot exclude the interactions of one electron with itself resulting in self-interaction errors, which can lead to wrong HOMO-LUMO gaps or an artificial charge delocalisation. On the contrary, in HF theory the self-interaction is cancelled by an exchange term. Due to that, hybrid functionals include a percentage of exact HF exchange to tackle this by reducing the self-interaction.

The included HF exchange is especially required for long-range interactions, but in hybrid functionals the exchange is scaled by a constant factor, α, (dependent on the functional, B3LYP $\alpha = 0.25$ HF exchange[88–90]). Hence, another improvement is established by long-range corrected (LC) hybrid functionals, which contain a separation into long and short-ranged Coulomb interactions and the HF exchange is just used for the long-range case. Table 3.1 shows the most commonly used functionals of the mentioned different classes, which were also used in this work.

Table 3.1.: Summary of different classes of exchange-correlation DFT functionals.

class	Approach	popular representation
LDA	E_{xc} depends just on the local ρ	SPW92[91]
GGA	E_{xc} depends on the local ρ and	PBE[92]
	on the local gradient ($\triangledown\rho$)	BLYP[88,93]
hybrid	E_{xc} also takes HF exchange into account	B3LYP[88,89,94,95]
LC	E_{xc} includes distance dependent HF exchange	$\omega B97xD$[96]

Kohn and Sham showed the existence of an electron density ρ, which is able to minimise the energy of an interacting many-body system and can also minimise a non-interacting (single-body) system in a defined external potential $v_{\mathrm{eff}}(r)$.[87] This leads to the well-known Kohn-Sham equation (Eq. 3.9),

$$\left[-\frac{1}{2}\triangledown^2 + v_{\mathrm{eff}}\right]\phi_i(r) = \varepsilon_i\phi_i(r) \tag{3.9}$$

resulting in an energy eigenvalue, ε_i, for the i-th electron in ϕ_i and a corresponding effective potential is $v_{\mathrm{eff}}(r)$, containing the following parts:

$$\begin{aligned} v_{\mathrm{eff}}(r) &= v_{\text{el-nuc}}(r) + v_{\text{el-el}}(r) + v_{\mathrm{xc}}(r) \\ v_{\text{el-nuc}} &= \frac{\delta E_{\text{el-nuc}}[\rho]}{\delta\rho(r)} = -\sum_a^N \frac{Z_a}{|r_i - R_a|} \\ v_{\text{el-el}} &= \frac{\delta J[\rho]}{\delta\rho(r)} = \int \frac{(\rho')}{|r_i - r'|}dr' + V_{\mathrm{xc}} \\ v_{\mathrm{xc}} &= \frac{\delta E_{\mathrm{xc}}[\rho]}{\delta\rho(r)} \end{aligned} \tag{3.10}$$

Further, the linear combination of atomic orbitals (LCAO) approach allows to represent the introduced MO in atomic basis functions $\eta(r)$, depending on their coefficients, c_μ^i

$$\phi = \sum_\mu c_\mu^i \eta_\mu(r) \tag{3.11}$$

Therefore, the Kohn-Sham Equation, Eq. 3.9, can be formulated in a more intuitive expression by introducing an Hamiltonian, $\hat{H} = \hat{T} + v_{\mathrm{eff}}(r)$.

$$\sum_\mu c_\mu^i \hat{H}[\rho]|\eta_\mu\rangle = \varepsilon\sum_\mu c_\mu^i|\eta_\mu\rangle \tag{3.12}$$

Eq. 3.12 can be simplified by multiplication from the left with $\langle\eta_\nu|$ and becomes

$$\sum_\mu c_\mu^i \langle\eta_\nu|\hat{H}[\rho]|\eta_\mu\rangle = \varepsilon\sum_\mu c_\mu^i\langle\eta_\nu|\eta_\mu\rangle \tag{3.13}$$

Furthermore, a matrix notation (see Eq. 3.14) can be used which will yield the orbital energies ε_i, overlap matrix, **S**, and orbital coefficients c_μ^i in **C**.

$$\mathbf{HC} = \mathbf{SC}\varepsilon \tag{3.14}$$

The electron density ρ can be calculated by using the occupied Kohn-Sham MO's, $\rho(r) = \sum_i \langle\phi_i|\phi_i\rangle$. Due to the case that $v_{\text{el-el}}$ and v_{ex} depend on $\phi_i(r)$, just an iterative solution is possible until it is self-consistent. The electronic energy of the system can then be calculated as shown in Eq. 3.15.

$$E[\rho] = \sum_i \varepsilon_i - \frac{1}{2}\int\int^{'} \frac{\rho(r)\rho(r')}{|r-r'|} dr dr' - \int v_{\text{xc}}(r)\rho(r)dr + E_{\text{xc}}[\rho] \tag{3.15}$$

The energy values of the HOMO or LUMO, $\varepsilon_{\text{HOMO/LUMO}}$, according to Janak's theorem[97], are directly related to the vertical ionisation potential (IP) or electron affinity (EA), respectively. The homologue holds for HF theory known as Koopman's theorem. Based on the unknown exact exchange-correlation functional, E_{xc}, a difference between the measured IP/EA and the orbital energy arises. For LDA and GGA a difference of $\sim$2 eV is reported, hybrids reduce this difference more and the best choice are the mentioned LC functionals[96,98] to reproduce the IP or EA.

Systems in which more than ~ 100 atoms are calculated by a QM method such as DFT, dynamical processes beyond the picosecond regime cannot be explored with an acceptable computational effort. Thus, e.g. chemical reactions can just be determined by single structure calculations to estimate the reaction time with respect to the energy barrier heigh. Then, it is not averaged over enough sampled structures and may under or over estimate the barrier height. Furthermore, the inclusion of the environment cannot be aligned with the limited system size.
Therefore, cheaper QM methods and further approaches may be required to allow the study of e.g. biological relevant systems in the nano second time regime. Also, thousands of calculations are required to obtain an averaged value for properties of interest from larger molecules. Therefore, semi-empirical methods, parametrized using DFT, allow to simulate the required time regime of nano seconds and systems of around 1000 atoms in an exactable computational time.

3.3. Density Functional Tight-Binding – DFTB

For many chemical problems, DFT yields really good results and is computationally more efficient compared to most wave function based quantum methods. Nevertheless, the mentioned problems in the time regime and system size limit the application of this method to larger systems.

To overcome this problem, semi-empirical and approximative methods are established,

which require less computational time and still produce quite accurate results. This is enabled by introducing fitted parameters. However, the transferability of the method is reduced which is based on the dependency between the used parameters and the system, used for fitting.

Density-functional tight-binding (DFTB) is parametrised from DFT and improves the computational speed by around 2-3 orders of magnitude, enabling to simulate larger molecules in sufficiently long QM/MM simulations.[99] Tight-Binding theories have their roots in solid state physics and are based on the assumption that the electrons are closely bound to the nuclei. The electronic density ρ can be obtained by using a reference density ρ_0 and its fluctuations.

$$\rho(r) = \rho_0(r) + \delta\rho_0(r) \tag{3.16}$$

Assuming this, the shown total energy of DFT, Eq. 3.7, can be expanded around the constructed electronic density of the atoms to yield the total energy for DFTB and the E_{xc} is expanded in a Taylor series up to the third order around the reference density $\rho_0(r)$

(3.17)

$$\begin{aligned}
E^{\mathrm{DFTB}}[\rho_0(r)+\delta\rho_0(r)] =& E_{\mathrm{xc}}[\rho_0(r)] + E_{\text{nuc-nuc}} - \int v_{\mathrm{xc}}[\rho_0(r)]\rho_0(r)dr - \frac{1}{2}\int\int^{'}\frac{\rho_0(r)\rho_0(r')}{|r-r'|}drdr' \\
&+ \sum_i \langle\phi_i| -\frac{1}{2}\nabla + v_{\mathrm{eff}} + \int\frac{\rho_0(r')}{|r-r'|}dr' + V_{xc}[\rho_0]|\phi_i\rangle \\
&+ \frac{1}{2}\int\int\left(\frac{\delta^2 E_{XC}[\rho_0(r)]}{\delta\rho(r)\delta\rho(r')} + \frac{1}{|r-r'|}\right)_{\rho_0(r)}\delta\rho(r)\delta\rho(r')drdr' \\
&+ \frac{1}{6}\int\int\int\left(\frac{\delta^3 E_{XC}[\rho_0(r)]}{\delta\rho(r)\delta\rho(r')\delta\rho(r'')}\right)_{\rho_0(r)}\delta\rho(r)\delta\rho(r')\delta\rho(r'')drdr'dr'' \\
=& E^0[\rho_0(r) + E^{1\mathrm{st}}[\rho_0(r),\delta\rho(r)] + E^{2\mathrm{nd}}\left[(\rho_0(r),\delta\rho(r))^2\right] + E^{3\mathrm{rd}}[(\rho_0(r),\delta\rho(r))^3] \\
=& E^{\mathrm{REP}} + \sum_i\langle\phi_i|\hat{H}[\rho_0(r)]|\phi_i\rangle + E^{2\mathrm{nd}}[(\rho_0(r),\delta\rho(r))^2] + E^{3\mathrm{rd}}[(\rho_0(r),\delta\rho(r))^3]
\end{aligned}$$

When the Taylor expansion is truncated after the first order, $E^{1\mathrm{st}}[\rho_0(r),\delta\rho(r)]$, the non-selfconsistent DFTB[100–102] is achieved, useful for solid state systems, see Eq. 3.22.

$$E^{1\mathrm{st}}[\rho_0(r)] = \sum_i\langle\phi_i|\hat{H}[\rho_0(r)]|\phi_i\rangle = \sum_i^{occ}\sum_\nu\sum_{mu} c_\nu^i c_\mu^i \hat{H}^0_{\nu\mu} \tag{3.18}$$

This term depends only on the reference density and is obtained by an atomic orbital Hamiltonian and just the valence electrons are taken into account while the core electrons are taken into account in the parametrisation of the $E^{\mathrm{REP}}[\rho_0]$ part.
The repulsive energy term, $E^{\mathrm{REP}}[\rho_0]$ depends on reference density ρ_0, is approximated by a sum of two-center calculations to reproduce DFT or empirical values and is also called $E^0[\rho_0]$. Three-center contributions are neglected and just the distance, $r_{\alpha\beta}$ be-

tween atom α and β are taken into account.

$$E^0[\rho_0] \approx E^{\mathrm{REP}}[\rho_0] = \frac{1}{2}\sum_{\alpha\beta} V_{\alpha\beta}^{\mathrm{REP}}[\rho_\alpha \rho_\beta, r_{\alpha\beta}] \tag{3.19}$$

This method is based on the tight-binding model, where electrons are expected to be enclosed to the nuclei, and therefore describable by a minimal basis, μ. Atomic orbitals (AO), ν, are used which are expressed by one single Slater-type function, $\eta_\nu(r)$, to reduce the amount of used basis functions to the minimum (minimal basis set approach). AO are a good choice to describe free atoms, but they are way to diffuse which leads to an inaccurate description of bonded chemical systems. Hence, a harmonic potential was applied by Eschrig[103] to overcome this problem, which dependents on an empirical parameter r_0, which is twice the covalent radius of the atom.

$$\left[-\frac{1}{2}\nabla^2 + v_{eff}(\rho_{\mathrm{atom}})(\frac{r}{r_0})^2\right]\eta_\nu(r) = \varepsilon_\nu \eta_\nu(r) \tag{3.20}$$

One large advantage of DFTB is that the diagonal elements of the Hamiltonian, $H^0_{\mu\mu}$, are precalculated and stored in tables, using DFT with the PBE[92] functional (LDA exchange-correlation functional see Table 3.1 for more informations). The off-diagonal elements are calculated using a two-center approximation, which relies on an atomic orbital μ located at atom α and on ν located at β. The values for the Hamiltonian and overlap matrix are tabled for distances until they are neglectable small. This speeds up the simulation in comparison to e.g. DFT where those integrals are directly calculated. However, this limits the transferability because the calculation of the atom-atom interactions can be performed for the tabled atoms only, requiring large parametrisations.

$$\begin{aligned} \hat{H}^0_{\mu\mu} &= \langle\mu|\hat{H}^0|\mu\rangle \\ \hat{H}^0_{\mu\nu} &= \langle\mu|\hat{H}^0|\nu\rangle = \langle\mu|\hat{H}^0[\rho_\alpha(r)+\rho_\beta(r)]|\nu\rangle \end{aligned} \tag{3.21}$$

Taken together, the total energy of DFTB1 can be expressed in terms of the repulsive term and the sum over all the eigenstates of the hamiltonian

$$E^{\mathrm{DFTB1}} = \frac{1}{2}\sum_{\alpha\beta} V^{\mathrm{rep}}_{\alpha\beta} + \sum_i^{occ} \varepsilon_i \ . \tag{3.22}$$

DFTB1 can calculate systems having an uniform charge distribution accurately and on the contrary it also performs good for systems having a complete charge separation such as NaCL.[104]

Taking the second order into account, called DFTB2, charge fluctuations can be considered which are important for the correct calculation of molecules.[105] This can be archived by taking the density fluctuation, $\delta\rho$, into account which can be written as a superposition of atomic contributions $\delta\rho = \sum_\alpha \delta\rho_\alpha$ and applying the monopole approx-

imation $\delta\rho \approx \Delta q_\alpha F_{00}^\alpha Y_{00}$. Then, it is assumed that the charge Δq_α is centred at the nucleus position, r_α, having a spherical shape such as a 1s-orbital, which allows to write the second order energy like the following

$$E^{\text{2nd}} \approx E^y = \frac{1}{2}\sum_{\alpha\beta}\Delta q\alpha\Delta q\beta\gamma_{\alpha\beta} \ . \tag{3.23}$$

An analytical function $\gamma_{\alpha\beta}$ is used to interpolate between two cases, shown in Eq. 3.24: a large distance between atoms α and β ($r_{\alpha\beta} \to \infty$) will result in a Coulomb-like description of the interaction between Δq_α and Δq_β.
A small distance ($\alpha = \beta$ and $r_{\alpha\beta} \to 0$) causes the Coulomb-like interaction to vanish and a resulting electron-electron repulsion at atom α rises, which is described by the Hubbard parameter, U_α. The Hubbard is approximatively twice the chemical hardness, and related to the size of the atom, $U_\alpha = 2\eta_\alpha$, defining the energy change of a system when an electron is added or removed.

$$\gamma_{\alpha\beta} = \begin{cases} \frac{\Delta q_\alpha \Delta q_\beta}{r_{\alpha\beta}} & if \ \ r_{\alpha\beta} \to \infty \\ \\ \frac{\delta^2 E_\alpha}{\delta^2 q_\alpha} = U_\alpha & if \ \ r_{\alpha\beta} \to 0 \end{cases} \tag{3.24}$$

This allows to write the total energy term of DFTB2 as the following:

$$E^{\text{DFTB2}} = \frac{1}{2}\sum_{\alpha\beta} V_{\alpha\beta}^{rep} + \sum_i^{occ}\varepsilon_i + \frac{1}{2}\sum_{\alpha\beta}\Delta q_\alpha \Delta q_\beta \gamma_{\alpha\beta} \tag{3.25}$$

DFTB2 can handle polar, organic systems and biomolecules quite good, however, the monopole approximation still introduces some problematic cases: the spherical shape of the charge density cannot take effects based on multipole moments into account such as the interactions of ion pairs. Furthermore, the used Hubbard parameter is independent of the charge state and performs bad for hydrogen atoms and therefore a further correction, $\gamma_{\alpha\beta}^h$, is added.

To overcome the mentioned failures the third-order terms can be included, DFTB3. In this method, derivatives of Hubbard are used to account for charge dependencies, thus, charge accumulation around an atom, α, increase or decrease the effective size of the atom. The derivatives of Hubbard are also precalculated and listed, which just increases the accuracy and not the calculation time sizeable. Therefore, the function Γ presents

the derivatives of Hubbard in the energy equation.

$$
\begin{aligned}
E^{3\mathrm{rd}} &= \frac{1}{3}\sum_{\alpha\beta}(\Delta q_\alpha)^2 \Delta q_\beta \Gamma_{\alpha\beta} \\
\text{with } \Gamma_{\alpha\beta} &= \left(\frac{\partial \gamma_{\alpha\beta}}{q_\alpha}\right)_{q_\alpha^0} = \left(\frac{\partial \gamma_{\alpha\beta}}{U_\alpha}\frac{\partial U_\alpha}{q_\alpha}\right)_{q_\alpha^0} \\
\text{with } \Gamma_{\alpha\alpha} &= \left(\frac{\partial \gamma_{\alpha\alpha}}{q_\alpha}\right)_{q_\alpha^0} = \left(\frac{\partial \gamma_{\alpha\alpha}}{U_\alpha}\frac{\partial U_\alpha}{q_\alpha}\right)_{q_\alpha^0}
\end{aligned}
\tag{3.26}
$$

Including the third order, DFTB3, improves the description of highly charged systems further.[99,106,107]

4. Molecular Mechanics

Ab initio quantum chemical methods can be used to calculate small systems ($N \approx 100$ atoms) in an acceptable time due to their scaling with N^3 or higher. Also, semi-empirical quantum methods are limited to several thousands of atoms, as shown in Figure 3.1. Based on that, larger systems with more than 1000 atoms cannot be studies by those methods. Additionally, the lengths of simulations of empirical and semi-empirical QM methods is limited to at least the nano second regime.
To study for example the structural reorientations of biological systems, such as proteins or DNA, different methods are requested. Therefore, molecular mechanics (MM) can be used to investigate larger systems ($N \geq 100k$ atoms) on timescales of more than microseconds. In this method, several approaches are applied to simplify the system: the electronic structure is reproduced by point charges, localised at spheres which represent the atoms and the interaction between atoms is described by a force field (FF) and depends just on their coordinates, R_N.

FF contains a sum of empirically parametrised potential energy functions, $V_{\mathrm{MM}}^{\mathrm{tot}}$, to calculate the interactions between atoms on a classical level. The bonds between atoms are pre-defined and treated as springs. Thus, chemical reactions cannot be taken into account. To allow a more detailed description of the same atoms, reproducing an individual chemical environment, atom types are introduced. Hence, several sets of parameters are listed to describe a carbon atom differently, e.g. a C-C bond in an aromatic cycle or in a protein backbone by applying different partial charges, q_i and bond lengths.

Taken together, the potentials and atom types are summarised as a so-called force field, which is optimised for a specific task such as the simulations of proteins. Many FF exist, differing in their reference data used to fit the potentials, their number of applied potentials and their individual definition of atom types (e.g. values of q_i). In this work AMBER force fields (AMBER-SB-ILLDN99* and GAFF) are used, containing the potentials described in the following. In general, the potential, $V_{\mathrm{MM}}^{\mathrm{tot}}$, can be divided into two groups, bonded and non-bonded interactions, shown in Eq. 4.1.

$$V_{\mathrm{MM}}^{\mathrm{tot}}(R_N) = V_{\mathrm{bonded}}(R_N) + V_{\text{non-bonded}}(R_N) \tag{4.1}$$

The bonded interactions, V_{bonded}, can be further subdivided to describe the interactions between all covalently bonded atoms with respect to their individual bond lengths, angels and dihedrals, shown in Eq. 4.2 and visualised in Figure 4.1. The predefined bond

between atom i and j is described by a harmonic potential, which is centred at a defined equilibrium distance r_{ij}^0 and has a force constant k_{ij}^b, to model the stiffness of the bond. The variation of the angle β_{ijk}, connecting three covalently linked atoms i, j and k, is constructed in the same way. The energy of the dihedral θ_{ijkl} is represented by a periodic potential, which describes the rotation along a covalent bond between j and k. Therefore, the dihedral describes the interaction between two plains constructed out of the atoms i,j,k and j,k,l, having the periodicity n and defined force constant V_n

$$\begin{aligned} V_{\text{bonded}}(R_N) = &\frac{1}{2}\sum_{bond}^{N_{bond}} k_{ij}^b (r_{ij} - r_{ij}^0)^2 \\ &+ \frac{1}{2}\sum_{angel}^{N_{angel}} k_{ijk}^a (\beta_{ijk} - \beta_{ijk}^0)^2 \\ &+ \frac{1}{2}\sum_{dihedral}^{N_{dihedral}} k_{ijkl}^d (1 + \cos(n \cdot \theta_{ijkl} - \theta_{ijkl}^0)) \end{aligned} \tag{4.2}$$

The non-bonded term, shown in Eq. 4.3, represents all non-covalent interactions between the N atoms of the system, which is modelled by the coulomb potential, representing the electrostatic interactions, and the Lennard-Jones potential, representing the van der Waals interactions and Pauli-repulsion.
Those interactions are considered for atoms, which are separated by more than four covalent bonds and the parametrisation of the bonded potentials take care of those interactions for atoms being separated by less than four atoms. All the mentioned potentials are visualised in a schematic way in Figure 4.1.

$$\begin{aligned} V_{\text{non-bonded}}(R_N) &= V_{\text{LJ}}(R_N) V_{\text{Coul}}(R_N) \\ &= \sum_{i \neq j}^{N} 4\varepsilon_{LJ} \left(\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right) + \sum_{i \neq j}^{N} \frac{q_i q_j}{4\pi\varepsilon_R \varepsilon_0 r_{ij}} \end{aligned} \tag{4.3}$$

The electrostatic interaction is based on the presented point charges $q_{i,j}$, located at the atoms i and j, and ε_0 is the vacuum and ε_R relative permittivity, respectively.
The Lennard-Jones potential depends on the distance, r_{ij}, between two atoms, ε_{LJ} describes the depth of the potential and σ_{ij}, defines the distance where the interaction is zero. The first part with the distance dependency $\sim \frac{1}{r^{-12}}$ describes the short-ranged Pauli repulsion, based on overlaps between orbitals, and the second belongs to the attractive, longer-ranged dispersion interactions $\sim \frac{1}{r^{-6}}$.
The calculation of the non-bonded interactions requires the largest computational effort, especially the calculation of the electrostatic interactions due to their larger long-range scaling $\sim \frac{1}{r_{ij}}$, while the Lennard-Jones potential can be neglected after distances of around ~ 10 Å and more.

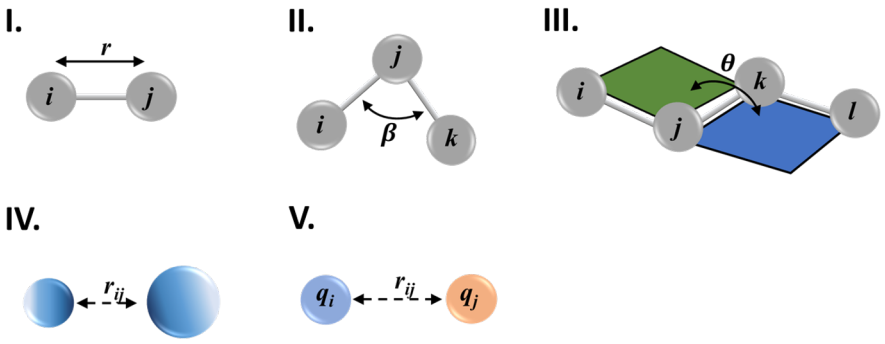

Figure 4.1.: Schematic view of the included FF potentials: **I.** shows the bond streching, **II.** the angle bending, **III.** the dihedral rotation, **IV.** van der Waals interactions and **V.** coulomb interactions

4.1. Molecular Dynamics Simulation

Normally, molecular mechanics methods are not used to calculate e.g. static properties of an optimised single structure of the system, such as quantum mechanical methods. The calculation of a property along a time series of structures, which is called a trajectory of the system, is used to generate an ensemble average, $\langle A \rangle_e$. This requires a correct sampling of the phase-space to ensure the ergodicity[108–110] of the system, which requires that the time-average and the average over the statistical ensemble are the same.

$$\langle A \rangle_t = \langle A \rangle_e \tag{4.4}$$

Therefore, a trajectory of the system is necessary which can be obtained by propagating the nuclear degrees of freedom using molecular dynamics (MD) simulations. Classical mechanics is considered to calculate the time evolution of the system by solving Newton's equation of motion (Eq. 4.5).

$$F_i = m_i \frac{d^2 R_i}{dt^2} \tag{4.5}$$

Thus, the size of the system N, the initial positions, R_i, of the i-atoms and their velocities, v_i, need to be defined. Then, the new positions and velocities are calculated iteratively *via* algorithms, such as the *leap-frog*[111] integrator shown in Eq. 4.6, which is used in this work.

$$\begin{aligned} v(t+\frac{1}{2}\Delta t) &= v(t-\frac{1}{2}\Delta t) + \frac{\Delta t}{m} F(t) \\ R(t+\Delta t) &= R(t) + \Delta t\, v(t+\frac{1}{2}\Delta t) \end{aligned} \tag{4.6}$$

$F(t)$, which is the requested force to propagate the i atoms, is the negative derivative of the total potential energy $V_{\mathrm{MM}}^{\mathrm{tot}}$ at the position R_i (shown in Eq. 4.7). Alternatively,

electronic structure calculations could also be used to calculate the force, called *ab initio* MD (possible for smaller systems only).

$$F_N = -\frac{\partial V_{\mathrm{MM}}^{\mathrm{tot}}}{\partial R_N} \tag{4.7}$$

The time step, Δt, between two iterations needs to be chosen carefully. The fastest motion of the represented system dominates this quantity, which corresponds to the vibration of hydrogen atoms, happening in $\sim$ 10 fs. Therefore, the time step is set to 1 fs. Applying algorithms constraining the H-bonds allows an increase of the time step to 2 fs, which was done in this work for the classical MD simulations, even the reduction to 0.1 fs needs sometimes to be applied to sample CT processes in mixed-quantum MD charge transfer simulations.
This propagation forms a trajectory representing the microcanonical ensemble, *NVE*, having a constant energy, system size and volume. However, the aim is a comparison to experimentally measured values, which are measured under a defined temperature T and pressure p. Therefore, the MD simulation needs to sample the *NPT* or isothermal-isobaric ensemble, which is ensured by introducing additional algorithms.
The temperature of the system is obtained by a thermostat regulating the velocities, such as the Nosé-Hoover thermostat,[112] using an additional degree of freedom to introduce a heat bath as internal part of the system, which can be seen as a friction term. The pressure coupling by a barostat, such as the Parinello-Rahman barostat,[113] can be done similarly, affecting the volume of the system. These algorithms create an oscillating temperature/pressure around the expectation value, and enable the sampling of a canonical ensemble. Thus, both algorithms are applied in the simulations of this work. Alternatively, the Berendsen thermostat/barostat can be applied, which simply scales the velocities or volume, using an exponential damping, to achieve a constant temperature or pressure but does not sample a canonical ensemble,

4.2. Quantum Mechanics/Molecular Mechanics Simulation

In the previous section the basics of QM and MM methods have been described and their individual strengths and limitations are discussed. However, both methods cannot be used to study chemical reactions in proteins, such as proton or electron transfers. Therefore, the straight-forward solution is a combination of both to use their individual strengths which is called a hybrid QM/MM approach.

The combination of QM and MM methods, requires that the system is divided into two subsystems, a QM and MM region. For example, the active site of an enzyme has to be calculated on a QM level, to allow bond formation/breaking due to their function which normally just involves a few amino acids. The remaining protein and solvent can be calculated on a classical MM level. Therefore, the Hamiltonian, $\hat{H}_{\mathrm{system}}$, has to be

divided into three parts, shown in Eq. 4.8, containing the pure MM, $\hat{H}_{\mathrm{MM}}$, pure QM, $\hat{H}_{\mathrm{QM}}$, and the interactions between both regions, which is the most challenging part, QM/MM, $\hat{H}_{\mathrm{QM/MM}}$.

$$\hat{H}_{\mathrm{system}} = \hat{H}_{\mathrm{QM}} + \hat{H}_{\mathrm{MM}} + \hat{H}_{\mathrm{QM/MM}} \tag{4.8}$$

To treat the interaction between both regions, several approaches were established and presented in the following:

- Mechanical embedding: this is computationally the simplest approach by just taking the electrostatic interaction between QM and MM region on a FF level, however, the QM region is not polarised by the MM region which can lead to an under- or overestimation of the interactions.
- Electrostatic embedding: the MM point charges enter the $\hat{H}_{\mathrm{QM}}$ as one-electron terms. Thus, the QM region gets polarised by updating the positions of the charges, corresponding to the MM region. This embedding increases the calculation time due to the expansion of the QM Hamiltonian, but just slightly and therefore it was also used in this work. In the applied MM/DFTB simulations the interactions are calculated adding a distance dependent term to the $\mathrm{E}^{2\mathrm{nd}}$ (see Eq. 3.25) in which $Q_{A_{\mathrm{MM}}}$ are the point charges of the classical atoms.

$$\sum_{A_{\mathrm{MM}}} Q_{A_{\mathrm{MM}}} \left(\frac{1}{r_{\alpha A_{\mathrm{MM}}}} + \frac{1}{r_{\beta A_{\mathrm{MM}}}} \right) \tag{4.9}$$

- Polarisation embedding: in this scheme also the MM charges are polarised by the fluctuation of the QM charges and *vice versa*. Therefore, polarisable FF needs to be involved which is currently not the usual procedure. This requires the calculation of the charges until self-consistency for the MM and QM level and therefore it is much more expensive and less frequently used.

The other interactions between the QM and MM region are treated by the FF terms such as the LJ potential. In most the QM/MM studies of proteins, the QM region needs to be cut out of the protein structure, therefore, several covalent bonds need to be cut. The most prominent way, to ensure this separation, is the introduction of link atoms. The cleaved bond is saturated by the introduction of hydrogen atoms for just the QM calculation. The link atom is very close to the point charges of the MM atoms and could lead to an overpolarisation of the QM region. To overcome this, the divided frontier charge scheme (DIV) [114] is introduced, to perform a charge spreading of the surrounding MM point charges.

4.3. Enhanced Sampling

MD simulations just yield the right properties, A, when the mentioned ergodic theory, $\langle A \rangle_t = \langle A \rangle_e$, is ensured. Based on that, the whole 6N dimensional phase space needs to be sampled, which is constructed out of the $3r$ coordinates and $3p$ momenta of all included atoms.[108–110] This is assumed for an infinite MD simulation

$$\lim_{t\to\infty} \frac{1}{t} \int_{t_0}^{t_0+1} A(t)dt = \int\int A(r,p) \frac{1}{Q} e^{\frac{-E(r,p)}{k_B T}} dr dp \tag{4.10}$$

in which t is the simulation time, E the energy, and Q the system partition function. This points out the strengths of MD simulations, but it also shows the limitation. To ensure the correct sampling endless simulations would be required or it would be ensured that the whole space is sampled. The first case is not suitable but the second one can be achieved using approaches which are introduced in the following.

In free MD simulations the system will fluctuate and sample configurations which are close in energy and belong to the equilibrium or educt state. Thus, the states, having higher energy or being separated by an energy barrier, are not sampled adequately. A transition between state **I.**, educt state, to the energetically lower state **III.**, product state, will be sampled in a free MD, if the barrier connecting them is just a few $k_B T$. Figure 4.2 shows a schematic representation of the free energy profile between state **I.** and **III.**, separated by a higher barrier along a chosen reaction coordinate, ξ.

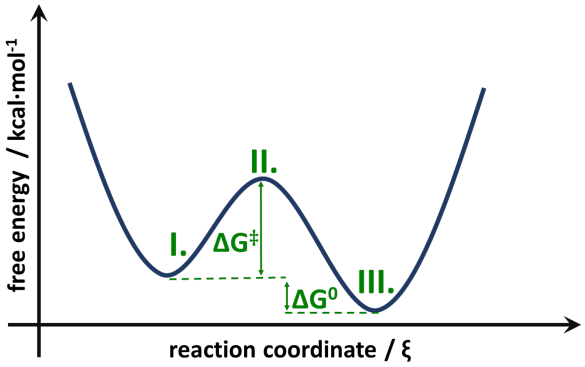

Figure 4.2.: Schematic view of a one dimensional free energy landscape: the population of the educt state **I.** is higher in energy than the product state **III.**.The transition state is presented at the point **II.**, having the highest energy. The barrier having the height of $\Delta G^{\ddagger}$ connects both states. The driving force which favours or disfavours a reaction is determined by the value of ΔG^0, ordering the character of the reaction as endergonic or exergonic.

The probability to be in state **I.** or **III.** depends on the energy, $P_A \sim \frac{-E(r,p)}{k_B T}$. Thus, states of higher energy will be undersampled or the transition to the product will never happen due to a high barrier. The free energy, F, exponentially depends on the energy of the sampled states (see Eq. 4.11). Therefore, the not adequately sampled phase space,

which is energetically higher and separated by a barrier, introduces large errors.

$$F = k_B T ln \left(\int \int e^{\frac{-E(r,p)}{k_B T}} \right) \rho(r,p) dr dp \tag{4.11}$$

On the other hand, properties like the height of the barrier, $\Delta G^{\ddagger}$, can be used to determine the reaction constant, k_{AB}, to identify the time regime. To solve this issue several approaches are obtained, called enhanced sampling methods. In the following, umbrella sampling and metadynamics will be introduced.

The selection of the reaction coordinate is a crucial source for errors, which is used to calculate the free energy. If the reaction involves degrees of freedom which are neglected in the calculation, the obtained free energy could never converge or it could under- or overestimate the correct energy profile. For example a direct transfer of a proton between a donor and acceptor can be simply modelled by one or two distances in a *in vacuo* simulation. However, in a protein pocket several other degrees of freedom could influence this transfer and need to be observed carefully. It could be required to included them into the reaction coordinate (or into the QM region). The gap of the free energy ΔG^0 between the educt and product state allows to determine the nature of the reaction and shows the yield or loss in free energy after the reaction to estimate the stabilisation of the product, **III.** or educt **I.** state.

4.3.1. Umbrella sampling

In Umbrella Sampling (US), a defined reaction coordinate, ξ, is required and then the change in free energy, F, along ξ is calculated. Hence, the free energy is defined as a function of ξ called potential of mean force (PMF):

$$F(\xi_0) = -k_B T ln P(\xi_0) + const.$$
$$\text{with}$$
$$P(\xi_0) = \frac{\int\int \delta[\xi - \xi_0] e^{\frac{-E(r,p)}{k_b T}} dr dp}{\int\int e^{\frac{-E(r,p)}{k_b T}} dr dp} \tag{4.12}$$

T represents the temperature of the system and $P(\xi_0)$ the probability to find the system in the interval $d\xi$ around ξ_0. To obtain the probability distribution, an integration over the whole phase space, excluding ξ, is required, which is done in MD simulations by obtaining the time average over the simulation. In principle, this can be achieved by just calculating the times in which the system was in the state ξ_0. However, this works just when the energy barrier is $\Delta G^{\ddagger} \sim k_B T$, otherwise the state will not be sampled adequate which will be the case for the transition state (TS) located at the barrier. This problem is tackled by US and an additional bias potential, $V(\xi) = \frac{1}{2} k(\xi - \xi_0)^2$, is introduced. Normally, a harmonic potential is used and located at selected ξ positions, constraining the system in that state to archive a correct sampling of that state. This approach is normally applied by performing separate simulations, which are called "Windows". The reaction coordinate is split and each individual window gets a defined value for ξ and an additional potential, $V(\xi)$, forcing the system to sample this configuration of ξ. The knowledge of the applied bias potential allows to calculate the free energy of the i-th window by the probability, P_i^{W}.

$$F(\xi) = -k_B T ln P_i^{\mathrm{W}}(\xi) - V_i(\xi) + \mathrm{const.}_i \tag{4.13}$$

The first part of Eq. 4.13 can be directly estimated by the simulation, the second is just the added known potential and the third one, $\mathrm{const.}_i$, is independent on ξ but unknown. To estimate the whole free energy landscape along ξ, all windows need to be analysed and to calculate the $\mathrm{const.}_i$ another algorithm such as the weighted histogram analysis method, *WHAM*, can be used. $\mathrm{const.}_i$ are vertical shifts of the free energy landscape which need to be excluded to obtain the right free energy curve, therefore, an overlap between neighbouring windows is required.

4.3.2. Metadynamics

Another biased sampling approach is the metadynamics simulation, which also applies additional potentials and requires a defined reaction coordinate, ξ, along which the

sampling is improved.[115] However, the added external potential is not designed in the same fashion as in the described US. Here, a history-dependent potential $V(\xi,t)$ is added to the Hamiltonian of the system. The bias potential is constructed during the simulation time: after a defined period, τ, a Gaussian is added to the current position, ξ. During the simulation the potential is a permanently extended sum of Gaussians. The shape of the Gaussians is predefined and can be modified to improve the sampling by changing the initial weight, σ_i, or height, ω.[116]

$$V(\xi,t) = \int_0^t dt' \omega e^{-\sum_{i=1}^{d} \frac{\xi i(r) - \xi i(r')^2}{2\sigma_i^2}} \tag{4.14}$$

This achieves penalisation of the current conformation by the Gaussian, because it is energetically not favoured any more and the system transfers into another configuration. Therefore, the picture of "filling" the free energy land scape by adding Gaussians enables to explore the whole phase space along ξ.

An improvement is the well-tempered metadynamics[117], in which the initial height of the Gaussian decreases in time to enable a better convergence of the free energy landscape

$$\omega = \omega_0 \tau e^{-\frac{V(\xi,t)}{k_B T}} \tag{4.15}$$

This results in a bias potential which smoothly converges to the underlying free energy profile which will result to a long time-limit,

$$V(\xi, t \to \infty) = -\frac{\Delta T}{T + \Delta T} F(\xi) + \text{const.} \quad . \tag{4.16}$$

The free energy, $F(\xi)$, is calculated along the reaction coordinate in a simulation, having the temperature T and a predefined parameter ΔT, called bias-factor $\gamma = \frac{\Delta T + T}{T}$ defining the maximum temperature range. Therefore, a bias factor, γ of 1 corresponds to "normal" MD simulation and for $\gamma \to \infty$ a "normal" metadynamic simulation and not a well-tempered one is performed.

This approach (and also US) is not limited to just one reaction coordinate, and therefore two-dimensional or higher free energy landscapes can be calculated. The convergence of the simulation allows a good measurement of the success of the sampling and the possibility to use the reaction coordinate to estimate the free energy.

5. Charge transfer

Charge transfer (CT) reactions play a central role in many biological systems to enable their functionality. In many essential processes and chemical reactions such as oxidative phosphorylation, photosynthesis or DNA repair, the transfer of an electron is the key role. Therefore, the simulation of this transfer allows a deeper insight to explore their mechanism at a molecular level. The calculations of those systems, especially the study of the mechanism of CT processes, need a complex methodology. On one hand the electronic structure needs to be represented for the CT, on the other hand the large systems cannot be treated completely by quantum dynamics methods. Even the small active regions for the CT cannot be sampled on a time scale at which the CT process happens. In biological systems two different regimes of CT are present and therefore discussed in the following: a long-ranged, superexchange tunnelling and a direct hopping CT mechanism.
In this section, the definition and construction of the energetic, which are used to describe the CT, will be explained. Then, the two different mechanisms of the transfer are presented. Afterwards, the concepts of the classical Marcus theory is summarised, which is a widely used approach to calculate the rates, k_{CT}, for the transfer in biological systems. Additionally, the limitations of Marcus theory and further developments to study a direct simulation, using a multi-scale simulation approach, of the CT will be presented.

5.1. Classical theories

In the following, an elementary CT event between a donor D and an acceptor A is considered, and a kinetic scheme of the process is shown in Eq. 5.1. Forward and backward transfer is possible, with respective rates k_{DA} and k_{AD}.

$$D^+ + A \underset{k_{DA}}{\overset{k_{AD}}{\rightleftarrows}} D + A^+ \tag{5.1}$$

This CT can be described as a transfer of a positive charge placed at the donor, initial state, and then transferred to the acceptor, final state. Therefore, it is called hole transfer due to the movement of the positive charge and schematically an ET in the opposite direction, which describes the same transfer process.

To describe a CT process, an adiabatic or diabatic picture can be used, which is displayed in Figure 5.1, showing the adiabatic states in black and the diabatic in the dashed green and blue lines. A natural choice of diabatic states is the charge being localised fully on the donor or acceptor, which will be called sites in the following. The choice of diabatic states is arbitrary, not unique and depends on the construction of the system, whereas the set of adiabatic states is unique. Hence, the diabatic representation can just be favoured in systems, allowing a clear construction of the diabatic states. The adiabatic energy, E_1, and diabatic energies for the donor, E_D, and acceptor, E_A, are identical as long as the charge is fully localised to either one of the sites. That changes in the transfer region around $Q^{\ddagger}$, however. While the diabatic states are degenerated here, the energies of the adiabatic states are split by twice the electronic coupling between these two states, $\pm H_{DA}$.

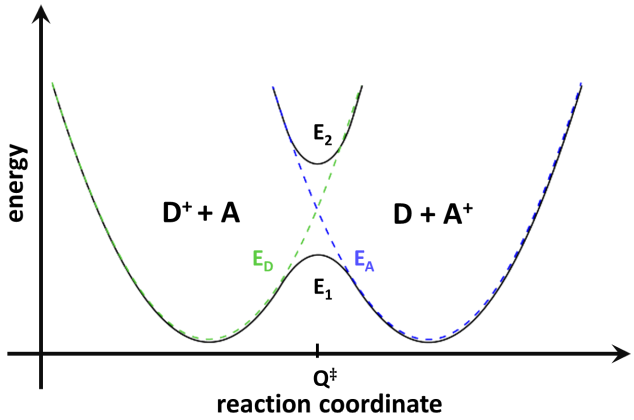

Figure 5.1.: Schematic construction of the Adiabatic and diabatic energy states: the adiabatic and diabatic energy landscapes along the reaction coordinate are shown in solid and dashed lines, respectively. The different diabatic states having the charge localised at the donor or acceptor are coloured in green or blue.

In the diabatic basis the energies E_D and E_A are the diagonal elements of the Hamiltonian and the couplings, H_{DA}, are the off-diagonal terms, shown in Eq. 5.2.

$$\hat{H}_{\text{diab}} = \begin{vmatrix} E_D & H_{DA} \\ H_{DA} & E_A \end{vmatrix} \tag{5.2}$$

It is assumed that the diabatic states are orthogonal and diagonalising of $\hat{H}_{\text{diab}}$, Eq. 5.2, gives the adiabatic states which correspond to the electronic ground and charge-transfer excited states, E_0 and E_1, respectively (shown in Figure 5.1).[15]

$$E_{0/1} = \frac{E_D + E_A}{2} \pm \frac{1}{2}\sqrt{(E_D - E_A)^2 + 4|H_{DA}^2|} \tag{5.3}$$

Eq. 5.3 shows the relation of the energies of diabatic and adiabatic states; the minus sign corresponds to the ground and the plus to the charge-transfer exited state.

Normally, the adiabatic states are electronically decoupled by applying the Born-Oppenheimer approximation. However, taking the nuclear Hamiltonian into account, the adiabatic states, ϕ_i and ϕ_j, are coupled *via* nuclear derivative couplings $\langle \phi_i(r;R)|\bigtriangledown_R|\phi_j(r;R)\rangle_r$. This is called the non-adiabatic coupling which can cause a transition between adiabatic states, whenever the velocity of the nuclei is large enough in the direction of the non-adiabatic coupling vector. Thus, including the motion of the nuclei and going beyond the Born-Oppenheimer approximation allows to consider the transitions between adiabatic states.

5.2. Tunnelling regime

The superexchange tunnelling mechanism is used to describes the transfer between an initial localised donor to a final localised acceptor using an intermediate structure as bridge, which happens on a time scale of up to milliseconds. A broad range of different structures are used as bridges, e.g. the backbone of the protein or non-covalently bond cofactors, and the donor/acceptor can also be cofactors, metal atoms or side chains of amino acids. The energy difference between the involved donor/acceptor and bridge structure can further divide the nature of the superexchange tunnelling process into a deep or resonant tunnelling. Hence, a CT involving an energy difference much greater than k_BT a deep tunnelling is assumed, otherwise a resonant tunnelling.[12–15] In deep tunnelling, the charge will have a very low probability to occupy the bridge directly and will tunnel through it directly.
On the contrary, in resonant tunnelling, the structural fluctuations can help to decrease the energy gap and the bride can be occupied, however, the residence time is not long enough that the surrounding structures of the bridge will respond to the new charge state of the system.

To study a superexchange tunnelling though, e.g. the aromatic ring of Phe or Tyr, a superexchange tunnelling calculation is applied. Therefore, the bridge needs to be included in the electronic coupling H_{DA} calculation. Hence, a definition of the site, describing the bridge B, is requested (selecting n orbitals of the bridge structure which are taken into account for the CT). The system is thus divided into the donor/acceptor (D/A) and bridge subspace and an effective Hamiltonian is calculated in which off-diagonal elements correspond to H_{DA} (for a detailed discussions see Ref. 118,119):

$$T_{DA} = \beta_{DA} + \sum_{i,j}^{B} \beta_{Di} G_{ij}(\varepsilon_{tun}) \beta_{jA} \tag{5.4}$$

here β_{DA} describes the direct electronic interaction between D and A, β_{Di} and β_{jA} the electronic interactions between donor, D, or acceptor, A, and the i or j-th orbital of the bridge. The tunnelling energy ε_{tun} corresponds to the average of the eigenvalues of the

effective Hamiltonian and G_{ij} is an element of the Green's function matrix, describing the probabilities for the electron to tunnel through the i-j space of the bridge.
On the contrary, if the charge stays longer after a transfer, the hopping regime can be used to describe the transfer. This holds for a CT along structures being energetically more similar, hence, thermal fluctuations enable the charge transfer. Normally, a transfer happens in the nano second regime and the distances for the transfer are shorter, compared to the tunnelled transfer, but continuous hopps can enable also a long-ranged ET.

5.3. Hopping regime

The CT can be described as a hopping event, assuming that the charge is localised on a molecule A or B e.g. the donor or acceptor, called site A or B, most of the time. The transfer of a positive charge from A^+ to B is assumed, see Scheme 5.1. In the hopping picture, the requested energy for a CT is higher than the energy for the relaxed state, therefore, thermal fluctuation of the molecule has to increase the energy of the sites, called site energy, which energetically allows then a transfer between the sites. Then, the molecule will relax again to form a resting state, which is lower in energy. A barrier rises due to that relaxation which prevents a direct continuous hop. Thus, thermal fluctuation needs to increase the energy again before the next hop can happen. Therefore, the charge transfer is a rare, infrequently event, however, the direct transfer between the sites is fast and the charge stays localised most of the time and is just shortly delocalised during the transfer.

If a harmonic response of the site energy fluctuations is assumed, the site energies can be described as parabolas, which represent the diabatic energy states (mentioned in section 5.1 part). Figure 5.2 shows the site energy parabolas and visualises the construction of important energetic quantities: the driving force ΔG^0 for a transfer between two sites, the diabatic activation energy $\Delta G^{\ddagger}$, which is the requested energy to move from site A and the transition state, and the reorganisation energy λ, which is the energetic cost to change the nuclear configuration of site A to the one of site B without changing the charge state.

The λ contains the two following parts: an inner-sphere relaxation, λ_i, taking the geometry change due to relaxation after switching the charge state. This can be calculated by four individual calculations, while the optimised geometries with respect to their charge states are calculated and the optimised geometries with the opposite charge state:

$$\lambda_i = [E_{A^+}(A) + E_B(B^+)] - [E_A(A) + E_{B^+}(B^+)] \tag{5.5}$$

The subscripts denote the charge state (hole on A or B) and the brackets show the charge state in which the geometry was optimised. The other contribution is the outer-

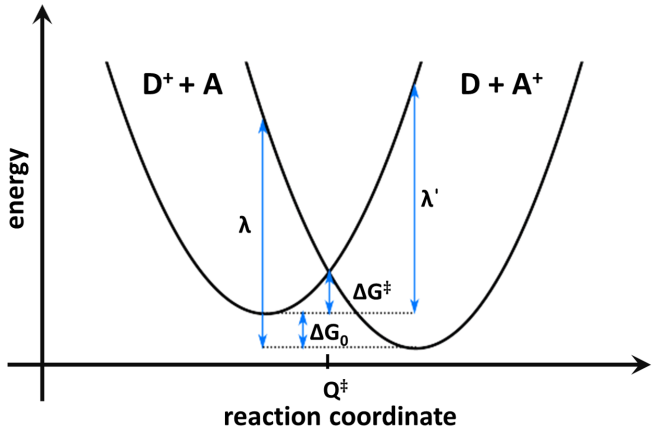

Figure 5.2.: Marcus theory parabolas: free energy profile according to Marcus theory for charge transfer from a donor to an acceptor. The free energy difference between both charge states is expressed by ΔG^0, and the reorganisation energy λ defines the curvature of the parabolas. The crossing point of both parabolas is the TS indicated as $Q^{\ddagger}$

sphere reorganization energy, λ_o, which describes the relaxation induced by rearrangements of the environment. Originally in Marcus theory, this was derived by the assumption of a surrounding polarisable continuum model, dependent on the transferred charge, Δq, the radii of the molecules $R_{A,B}$ and ε_{op}, ε_s which are the optical and dielectric environment constants, respectively.

$$\lambda_o = (\Delta q)^2 \left(\frac{1}{2R_A} + \frac{1}{2R_B} - \frac{1}{2R_{AB}} \right) \left(\frac{1}{\varepsilon_{op}} + \frac{1}{\varepsilon_s} \right) \tag{5.6}$$

However, the use of MD simulations allowed a direct sampling of the environment, therefore, a simulation is used to sample the states, having the charge at A and another one having the charge at B. The trajectories allow to calculate the averaged values and their difference is λ_o:

$$\lambda_o = \langle E^{\mathrm{MM}}_{AB^+} \rangle_{A^+B} - \langle E^{\mathrm{MM}}_{AB^+} \rangle_{AB^+} \tag{5.7}$$

Furthermore, the effect of the reorganisation energy can be considered directly in simulations of CT processes using a multi-scale approach, which will be introduced in section 5.5. In both cases, there is a problem with the missing description of electronic polarisability of the environment; this leads to an overestimation of the reorganisation energy, and thus an underestimation of the CT needs to be considered. The reorganisation energy and the whole picture is based on the parabolic description, therefore, the same curvature of both parabola is assumed which corresponds to the assumption that $\lambda = \lambda'$.

To study those transfers, the reaction rate, k_{AB}, is the value of interest, determining the time regime of the CT in e.g. the protein. Warshel and Zusman showed that the vertical

diabatic energy gap $\Delta E = E_A - E_D$ can be used as a suitable reaction coordinate for a CT.[120,121] In the harmonic approximation, semiclassical transition state (TS) theory can be used while the diabatic activation energy, $\Delta G^{\ddagger}$, is reduced by an adiabatic correction factor $\Delta^{\dagger}$:

$$k_{AB,\mathrm{TST}} = \nu_{\mathrm{eff}} \kappa_{\mathrm{el}} \exp \frac{\Delta G^{\ddagger} - \Delta^{\dagger}}{k_B T} \tag{5.8}$$

Here, ν_{eff} is the requested frequency of the effective mode of motion, which needs to be activated for the transfer to take place. κ_{el} is the electronic transmission coefficient, which describes the efficiency of the CT event once the transition state $Q^{\ddagger}$ has been reached. For low temperatures, a tunnelling correction can also be applied, lowering the energy barrier slightly.[122] Both prefactors of Eq. 5.8 may be obtained from Landau–Zener theory (LZ),[123–125] in which P_{LZ} is the probability for a single crossing of the TS along the reaction coordinate.

$$\kappa_{\mathrm{el}} = \begin{cases} \dfrac{2P_{LZ}}{1+P_{LZ}} & \text{if } \Delta G^{\dagger} \geq -\lambda \\ 2P_{LZ}(1-P_{LZ}) & \text{if } \Delta G^{\dagger} < -\lambda \end{cases} \tag{5.9}$$

$$P_{LZ} = 1 - \exp[-2\pi\gamma] \tag{5.10}$$

$$2\pi\gamma = \frac{\pi^{3/2} \langle |H_{AB}|^2 \rangle_{\mathrm{TS}}}{h \nu_{\mathrm{eff}} \sqrt{\lambda k_B T}} \tag{5.11}$$

The electronic coupling between the initial and final diabatic state $\langle |H_{AB}|^2 \rangle_{\mathrm{TS}}$ needs to be averaged over all nuclear configurations. It was shown that the two contributions of the activation energy $\Delta G^{\dagger}$ can be expressed within the LZ theory as[126]

$$\Delta G^{\ddagger} = \frac{(\lambda + \Delta G^0)^2}{4\lambda} \tag{5.12}$$

$$\Delta^{\ddagger} = \sqrt{\langle |\hat{H}_{AB}|^2 \rangle_{\mathrm{TS}}} - \frac{\langle |\hat{H}_{AB}|^2 \rangle_A}{\lambda} \tag{5.13}$$

The first part of the adiabatic correction factor expresses the averaged difference between the adiabatic and diabatic energy surfaces at the TS, and the second part the same energy difference averaged at the initial state.

The nature of the CT can be defined by an adiabatic or non-adiabatic limit, which is based on the behaviour of the system near the TS, $Q^{\ddagger}$. Due to that, the transfer may proceed slowly, also called *adiabatically*,[15] from the initial to the final state – this happens with large electronic couplings and small values of reorganisation energy, $H_{AB} > \lambda$. On the contrary in the non-adiabatic limit, the wave function changes nearly instantly after crossing the barrier, which is the case for small couplings and high reorganisation energies, $H_{AB} < \lambda$. This non-adiabatic limit makes a further simplification possible: assuming $\gamma \ll 1$, which corresponds to a very fast transfer at the TS, $e^{-2\pi\gamma}$ may be expanded to the first order, so that $1 - e^{-2\pi\gamma} \approx 2\pi\gamma$. Finally, this leads to the widely used

Marcus' expression for the transfer rate:[127,128]

$$k_{\mathrm{CT}} = \sqrt{\frac{\pi}{\lambda k_B T}} \left\langle |H_{AB}|^2 \right\rangle_{\mathrm{TS}} \exp\left[-\frac{\Delta G^{\ddagger}}{k_B T}\right] \tag{5.14}$$

5.4. Limitations of classical theories

Marcus theory was derived to describe CT processes in solvated metal complexes originally, but it became frequently used to also describe CT in biological systems[129] and organic semiconductors.[118,130] The extension of the computational power in combination with QM/MM approaches allowed the treatment of larger systems, such as biomolecules. An example is the CT between covalently bound heme cofactors in a tetra-heme cytochrome.[131] Thus, applying Marcus theory to calculate CT rates, k_{CT}, in biomolecular systems, having a clearly defined hopping transfer between fully localised charge states, showed successful reproductions of the experimental measurements in various studies. After the CT, the site is able to relax fully until the next transfer, and therefore the required equilibrium based mechanism of Marcus theory can be used.

However, sources such as a different environment can offer a weaker or stronger stabilisation, which can cause a different transport mechanism. This can result for example in a driving force of the transfer, guiding the charge into a direction or forcing a larger resting time at a specific site. The charge may not stay long enough after a hopping event, therefore, the relaxation and decrease of energy will not happen when the charge hops to another site too frequently. Strong couplings or driving forces may cause this behaviour which can be based on the environment of the transfer path. However, the energy gap between the two sites is not corresponding to a superexchange tunnelling transfer, in which a structure between a donor and acceptor is just populated shortly by the resting charge (described previously, see section 5.2).

For example, the enzymes Photolyases (PL) and signalling proteins Cryptochromes (CRY) show a conserved CT path which starts in the centre of the protein and ends at the surface, having an aqueous environment. Their pathway for the CT involves the same amino acids, having no significant energy gap, and can be described by three similar diabatic states. However, the polarisability of the environment around the three sites forms an increasing interaction with the charge. This allows a stronger stabilisation on the charge on the last site of the pathway. Therefore, a driving force enhances the transfer and directs it from the first to the last site. Furthermore, this decreases the resting time at the site in the middle of the path and does not enable a full relaxation. Thus, it was demonstrated that that the nature of the transfer is not based on an equilibrium process.[16]

Marcus theory can be used to reproduce the experimental reported transfer rate for

the first CT, however, fails to reproduce the second CT. Hence, a transfer rate of less than 0.0003 ns^{-1} is gained for the second CT in class-I CPD-PL of *E. coli*, while the experimental measurement observed a transfer in 33 ns^{-1}.[16] This is based on the fact that the charge will not remain long enough at this site to relax fully and therefore no barrier for the next transfer arises. However, assuming a Marcus picture for this transfer will yield a lager barrier because the second site relaxed fully and this corresponds to the underestimation of the rate by a factor of more than 10^4.
The same could also happen in systems of organic liquid crystals offering the possibility to hop fast along several sites due to a good orientation and resulting high coupling or small reorganisation energy. Hence, CT requires a different approach which involve non-equilibrium processes and is in a hopping-like regime. The charge carrier needs to interact directly with the environment, therefore, no averaged values for the coupling and energy gap can be used to calculate a transfer rate. Hence, the changing position of the charge causes too large variations. A solution of this is a direct coupled dynamics scheme, a non-adiabatic MD simulation, which is a fundamental factor in the study of those biological systems and was able to reproduce both CT rates for the PL system in agreement with the experiment. The details and assumed approximations of this approach will be described in the following section.

5.5. Coupled quantum – classical dynamics simulation

In several biomolecular complexes, the CT process cannot be described by a Marcus model. Therefore, alternative methods need to be used, allowing the study of e.g. hopping-like CT between not fully equilibrated states. The interaction between electronic and nuclear degrees of freedom needs to be considered explicitly to describe the time-dependent relaxation of atomic centres (nuclei).

However, the time-dependent SE for both cannot be solved in the molecular systems of interest due to their extensive sizes. An approach, which uses calculations of different levels, needs to be used, such as non-adiabatic semi-classical dynamics based on optimised quantum chemistry methods. Therefore, the dynamics of nuclei are calculated by classical mechanics (see section 4.1) and the transferred charge is treated by solving the time-dependent SE, which employs QM methods (see section 3.2) or semi-empirical methods such as DFTB (see section 3.3), which speeds up the simulation further and allows to simulate larger QM region.

In non-adiabatic semi-classical simulations schemes, the propagation of the nuclei is coupled to the propagation of the charge and *vice versa*, using a mean-field (Ehrenfest) or surface-hopping (SFH) approach. A non-adiabatic molecular dynamics (CT-QM/MM MD) using DFTB for the electronic structure calculation and a force field method for the remaining system was able to reproduce the experimental measured CT processes in

DNA,[132–135] in members of the PCF[16–18] and also in organic semiconducting materials.[130,136,137] This multi-scale method is used to obtain the results which are presented in chapter 6 and 7 to characterise the CT of several more members of the PCF and further extended in chapter 8 to study the impact of the propagator using test cases.

5.5.1. Fragment orbital approach – FO

Non-adiabatic CT-QM/MM MD requires the molecular system to be split into a MM and a QM region, and additionally, the QM region is further divided into individual fragments. This fragmentation requires systems which can be separated into several fragments without introducing unphysical effects. For instance, the Trp Triad in PCF, the nucleobases in DNA, or the individual molecules of crystalline organic semiconductors are convenient choices of fragments within this approach. Furthermore, such a separation based on the nature of the system offers a clear definition of the diabatic states, which correspond to a fully localised electron on each individual fragment. The bonds crossing the QM-MM boundary are saturated by placing hydrogen atoms as linking atoms. The electronic structure calculation is the most time-consuming step. Simulations in the nanosecond timescale, which are requested to characterise and study the biological CT in PCF and sufficient large systems to study CT in the organic crystals, cannot be achieved by *ab initio* DFT or HF calculations. That makes it necessary to introduce the following approximations that reduce the computational time and to further archive a linear scaling of the calculation time with respect to the system size: the semi-empirical DFTB method is employed for the QM calculations, and a Fragment-Orbital (FO) approach, is applied and therefore just selected orbitals of interest are considered (e.g. for hole transfer the HOMO and for electron transfer the LUMO).

In the FO approximation the diabatic states, Ψ_A and Ψ_B, can be constructed directly by choosing a set of orthogonalised orbitals which are localised at the different fragments. For example assuming a dimer system with two adiabatic states and also two fragments will be presented, one will carry the charge initially, corresponding to the donor, and the other will be occupied after the transfer, corresponding to the acceptor.
In the first step, individual DFTB calculations of isolated fragments yield two sets of orbitals $[\phi_A^i]$ and $[\phi_B^i]$. The orbitals within one set are orthogonal by construction, however, a small overlap between two sets occurs normally. Therefore, the Löwdin method[138] shall be used to generate a orthogonal set $[\phi_A^i;\phi_B^i]$. If the Hamiltonian is expressed by independent-particles such as in the Kohn-Sham DFT (see Eq. 3.7) the integral between both orbitals represents the electronic coupling between those orbitals, H_{AB}, which was mentioned in the previous section. Thus, the FO approach enables the electronic coupling calculation between two selected diabatic states.

$$\langle \phi_A^i;\phi_B^i\rangle \approx \langle \Psi_A|\hat{H}|\Psi_B\rangle = H_{AB} \tag{5.15}$$

5.5.2. Approximate total energy expression with the FO approach

The starting point for the electronic structure calculation is the energy of the charge-neutral closed-shell system with N electrons, the DFT ground state density ρ_0 and ϕ_0^i are Kohn–Sham orbitals. This approach follows the DFTB2 approach, see Section 3.3, and can also be mentioned as a coarse-grained DFTB2 in which the selected FO takes the role of the single atoms and the energy of the neutral system can be expanded up to the second order as in Eq. 3.17.
A frozen orbital approximation is applied, which assumes that $\Psi_0 = \Psi_{N-1}$, and the energy for a system having $N-1$ electrons can be obtained by an expansion of the energy, $E^N[\rho_0]$, as a function of the density $\rho = \rho_0 + \delta\rho$ around the density ρ_0 up to the second order $E^{2\mathrm{nd}}[\rho, \delta\rho]$.

$$E^{N-1}[\rho] \approx E^N[\rho_0] - \langle\Psi_0|\hat{H}[\rho_0|\Psi_0\rangle + E^{2\mathrm{nd}}[\rho_0, \delta\rho]. \tag{5.16}$$

The calculation of $E^N[\rho_0]$ by using MM methods, much like for the remainder of the system, allows a further reduction of the computational cost.[130]

$$E^N[\rho_0] = E^{\mathrm{MM}} \tag{5.17}$$

The applied FO approach constructs the wave function Ψ_0 as a linear combination of the selected orbitals (i is HOMO or LUMO) of the fragments, instead of using all the AO such as in DFTB.

$$\Psi_0 \approx \sum_M a_M^i \phi_M^i \tag{5.18}$$

Using this basis set expansion, the CT matrix elements can be calculated as

$$\Psi_0 = \sum_M a_M^i \phi_M^i \rightarrow \langle\Psi_0|\hat{H}[\rho_0]|\Psi_0\rangle = \sum_{MN} a_M^* a_N \langle\phi_M^i|\hat{H}[\rho_0]|\phi_N^i\rangle = \sum_{MN} a_M^* a_N H_{MN}^0 \tag{5.19}$$

The diagonal elements, $H_{MM} = \varepsilon_m$, are called site energies and the off-diagonal terms are the electronic couplings between the fragments, H_{MN}. Due to the similar construction as DFTB, this can be called a coarse-grained DFTB Hamiltonian as shown in Eq. 3.18. The variation of the density, $\delta\rho$, consists of contributions of each of the M fragments, $\delta\rho = \sum_M \delta\rho_M$, thus $E^{2\mathrm{nd}}$ can be constructed as

$$E^{2\mathrm{nd}} = \frac{1}{2}\iint' \left(\frac{1}{|r-r'|} + \frac{\delta^2 E_{xc}}{\delta\rho\delta\rho'}\Big|_{\rho=\rho_0}\right)\Delta\rho_M \Delta\rho_N' \mathrm{d}r\,\mathrm{d}r' \tag{5.20}$$

Similarly as in DFTB, a monopole approximation is used to express $\delta\rho_{0M}$

$$E^{2\mathrm{nd}} = \frac{1}{2}\sum_{M,N} \Delta Q_M \Delta Q_N \Gamma_{MN} \tag{5.21}$$

This simplifies Eq. 5.20, and the occupation of fragment M can be obtained as $\Delta Q_M = |a_M|^2$. Two different cases for Γ_{MN} are considered:

$$\Gamma_{MN} = \begin{cases} U_M & \text{for } M = N \\ \frac{1}{R_{MN}} & \text{for } M \neq N \end{cases} \tag{5.22}$$

Γ_{MM} is the Hubbard parameter, related to the chemical hardness, of fragment M, which is used to model the onsite electron-electron repulsion in the same fashion as the atomic Hubbard in DFTB. Γ_{MN} is used to model the Coulomb interaction between the charges localised at fragments M and N. The additional charge, with respect to the localisation of site M is then decomposed into an atom dependent charge $\Delta Q_M = \sum_{\alpha=1}^{N^{atoms}} \Delta q_\alpha$, therefore, the interaction between the atoms, m, of the charged fragment M and the environment can be calculated with respect to the structure of the site.

Since GGA functionals exhibit the notorious self-interaction error for radical systems due to the unpaired electron, the second-order term needs to be scaled down to compensate for this error. A scaling factor $C = \frac{1}{5}$ was established.[139] Taken together, the following total energy expression arises:

$$\begin{aligned} E^{N-1}[\rho] &\approx E^N[\rho_0] - \sum_{MN} a_M^* a_N \langle \phi_M^i | \hat{H}[\rho_0] | \phi_N^i \rangle + C \cdot E^{2\text{nd}} \\ &= E^{\text{MM}} - \sum_{MN} a_M^* a_N H_{MN}^0 + C \left(\frac{1}{2} \sum_M \sum_N \Delta Q_M \Delta Q_N \Gamma_{MN} \right) \end{aligned} \tag{5.23}$$

The variable atomic charges Δq_α, on the charge-carrying fragments, are updated in the charges presented in the MM simulation steps. Thus, the environment is directly affected by the position of the charge, which is the outer-sphere reorganization energy. The geometry of the fragment would also be changed with respect to the occupation, however, the bonded parameters are not updated to reproduce a change with respect to the varying charge state. Therefore, no adoption of the energy is included, which was described in section 5.3 as inner-sphere reorganisation energy.
This missing interaction is corrected by using a precalculated parameter, λ_i^M, which is added to the site energy of the fragments. A system consisting of different sites such as Tyr and Trp, needs another precalculated parameter, shifting the site energy, $\Delta_{\text{shift-DFT}}$. This is based on a well-known error of DFTB, which occurs by comparing absolute values of the HOMO energies. A small comparison with DFT is used to overcome this error, which determines the HOMO energy levels of different sites. Taken together, the site energy has the following contributions:

$$\varepsilon'_M = \varepsilon - \lambda_i^M \Delta Q_M^2 + \Delta_{\text{shift-DFT}} + C \sum_N \Delta Q_N \Gamma_{MN} \tag{5.24}$$

5.5.3. Simultaneous propagation of electronic and nuclear degrees of Freedom

The Lagrangian formalism is applied to the expression for total energy Eq. 5.23 to obtain the coupled equations of motion for the nuclei, described with MM, and the propagation of the charge carrier which involves to solve the electronic SE for a time-dependent electronic potential created by the classical nuclei. Therefore, the time-dependent electronic wave function, $\Psi(t)$, can be expressed as a linear combination of a sum of adiabatic, $\varphi_n(R(t))$, or diabatic electronic wave functions, $\phi_N(R(t))$, which depend on the nuclear coordinates, R,

$$\Psi(t) = \sum_n a_n(t)\varphi_n(R(t)) \ . \tag{5.25}$$

The combination of Eq. 5.25 and the time-dependent SE, $i\hbar\dot{\Psi} = \hat{H}\Psi$, yields a time-evolution of the expansion coefficients a_n

$$i\hbar\dot{a}_n(t) = \sum_n a_n(t)(H_{mn} - i\hbar d_{mn}) \ . \tag{5.26}$$

Here, H_{mn} are the matrix elements of the Hamiltonian in the adiabatic basis, $H_{mn} = \langle\varphi_m|H|\varphi_N\rangle = E_n\delta_{mn}$ and $d_{mn} = \langle\varphi_n|\frac{\partial}{\partial t}|\varphi_m\rangle$ are the non-adiabatic coupling matrix elements.

In the previously mentioned non-adiabatic CT-QM/MM MD scheme, enabling the simulation of an additional charge, the diabatic states can be defined straight forward and will be used. Therefore, the time-dependent electronic wave function, $\Psi(t)$, is expressed as a linear combination of a sum diabatic states, already defined in Eq. 5.18. The construction of the diabatic states, corresponding to localised charged states, which allows the assumption that $\langle\phi_N|\frac{\partial}{\partial t}|\phi_M\rangle \approx 0$.[130] Thus, simplifies Eq. 5.25 further and $-i\hbar d_{mn}$ can be neglected, which leads to the following description for the time-evolution of the expansion coefficients, a_N, in a diabatic basis (determined by the capital subscript).

$$i\hbar\frac{\partial a_M}{\partial t} = \sum_N a_N\langle\phi_M|H_{MN}|\phi_N\rangle = \sum_N a_N H_{MN} = \sum_{N\neq M} a_N T_{MN} + a_M\varepsilon_M \tag{5.27}$$

The advantage of this representation for the CT derives a connection between the charge population of site M with the coefficient a_M which is calculated directly between two steps of the classically propagated system.

$$Q_M = e|a_M|^2 \tag{5.28}$$

The classical charged atoms interact with the current position of the propagated charge,

therefore, additional forces are obtained.

$$m_N \ddot{R}_N = -\frac{\partial E^{\text{tot}}_{\text{MM}}(q^0_A, \Delta q_\alpha)}{\partial R_N} \tag{5.29}$$

In this non-adiabatic CT-QM/MM MDsimulation approach the motion of the nuclei is coupled to the propagated additional charge. Thus, the environment of the CT active molecular complex becomes polarised by the travel of the charge directly. The individual steps of this semi-classical approach are schematically shown in Figure 5.3, to address the procedure in which the different calculations are looped. Furthermore, different methods to propagate the charge carrier can be used in the semi-classical approach. Two, which are applied in this work, are introduced in the following briefly.

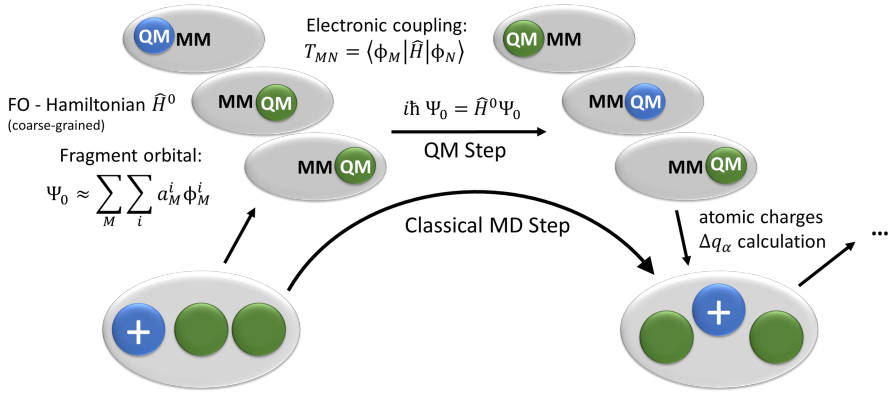

Figure 5.3.: Schematic view of CT-QM/MM MD: the lower, larger spheres indicate the system for the MM propagation and the smaller three spheres above the decomposed QM system (three sites). The position of the charge is highlighted by the blue sphere and the green ones are neutral sites. The left three spheres indicate the individual site calculations and then $\hat{H}^0$ is diagonalised, the coupling calculated and the new population of the sites are determined. Therefore, the charge moved from the first to the middle site. The left arrow between the QM and MM region represents the update of the MM charges until the next MD step will be performed, which is visualised by the three dots.

In the Mean-field (Ehrenfest) propagation, an averaged potential energy surface is used. Therefore, all adiabatic states are taken into account and weighted by their population to yield the averaged potential energy surface. Hence, the environment interacts with just a combined quantum state. Another propagation scheme is the surface hopping (SFH) method in which the QM system is also allowed to stay in different defined adiabatic states. In the SFH the populations of states are also propagated in time and the system occupies different adiabatic states and after a defined period of time, δt_{hop}, transitions to another state can occur. However, one adiabatic state is selected, the active surface, which affects the surrounding MM atoms. The calculation of the probability to "hop" into an other state is the core of the propagation. The difference to e.g. the Ehrenfest propagation arises in the different quantum forces applied to the surrounding nuclei. In SFH just the most populated surface is used to derive this force, while in the Ehrenfest propagation the averaged state is used for that. Furthermore, the following assumptions

are applied to ensure the correct propagation of the QM system:

- The probability, $P_{n\to m}$, to hop between two adiabatic surfaces, n and m, is calculated using Tully's well-known fewest switch algorithm,[140] in which the probability to hop between state m to n is defined as the following:

$$P_{n\to m} = \delta t_{hop} \frac{2Re\ a_m a_n^* d_{mn}}{a_m a_n^*} \quad . \tag{5.30}$$

  In the current implementation (used in chapter 8) a numerical approximation is used for d_{nm}:

$$d_{nm} = \langle \varphi_n | \frac{\partial}{\partial t} | \varphi_m \rangle \approx \langle \varphi_n(t - \delta t_{hop}) | \varphi_m(t) \rangle \quad . \tag{5.31}$$

- The decoherence correction introduced by Persico[141] is used to approach correct transfer rates. The expansion coefficients of the unoccupied surfaces are exponentially decayed, using the decoherence time proposed by Truhlar.[142]

$$a_m \leftarrow c_m e^{\frac{-\delta t_{hop}}{\tau_{nm}}} \tag{5.32}$$

$$\text{with } \tau_{nm} = \frac{1}{|E_n - E_m|} \left(1 + \frac{C}{E_{kin}}\right) \tag{5.33}$$

  This depends on the kinetic energy, E_{kin}, the energy eigenvalues of the states (E_n, E_m) and an adaptable parameter C.

- Additionally, a flexible SFH approach Ref. 143, which is introduced by Beljonne. This reduces the trivial crossings by simply reducing the size of the system. The adiabatic states m, n are derived out of diabatic states, which are directly constructed by the nature of the system. Thus, in this flexible approach not all diabatic states are taken into account to construct the adiabatic states. If the coupling between the current active surface and the diabatic state reaches a threshold, R_c, this diabatic state will be included for the next hopping step. Therefore, less adiabatic states arise this leads to less trivial crossings and the corresponding unphysical long-ranged CT. This is explicitly derived in Ref. 143 and just applied here in the same fashion as in Ref. 144

Part III.

Results

6. Charge transfer processes in a 6-4-Photolyase – *PhrB*

Reproduced in part from Ref. 26

(Functional role of an unusual tyrosin residue in the electron transfer chain of a prokaryotic (6-4) Photolyase)

with the permission of the Chem. Sci. Owner Societies.

Reproduced in part from Ref. 27

(Two aspartate residues close to the lesion binding site of Agrobacterium (6-4) photolyase are required for Mg2+ stimulation of DNA repair

with the permission of the FEBS Journal.

In *Agrobacterium tumefaciens* a 6-4 photolyasen-homologues protein, called *PhrB*, was identified. Phylogenetic studies by Lamparter and co-workers propose a close relation to the ancestral Photolyase (PL) which changed the hypothesis of the evolution of the PCF by introducing a new ancestor subgroup, the FeS-BPC (see Figure 1.1 and Ref. 24,25 for further descriptions). The structure of *PhrB* shows differences in the protein architecture, especially in the Triad which is essential for the charge transfer (CT) in the photo-activation. PCF normally have a conserved Triad,[19] consisting out of three tryptophan (Trp) residues, to enable CT between the surface of the protein and the excited FAD to form $FAD^{\bullet -}$or $FADH^{-}$(see section 1.1.3 for further descriptions). Tyrosine (Tyr) can also be involved in the CT in PCF, e.g. to offer an alternative pathway (CRYI *Arabidopsis thaliana*) or an extension of the Triad (*A. nidulans* PL). Normally, a direct proton acceptor is required because $Tyr^{\bullet +}$ is unstable and needs further deprotonation to form a stable intermediate. However, *PhrB* contains a Tyr391 and not Trp in the first position of the conserved Triad pattern, which is embedded in the centre of the protein

without an obvious proton donor. Trp342, located at the periphery, and Trp390, connecting Trp342 and Tyr391 or FAD are in the conserved positions and complete the construction of the Triad (Tyr391, Trp390 and Trp342 are abbreviated in this chapter as **A**, **B** and **C**, respectively). The protein structure of *PhrB* and the geometry of the Triad are illustrated in Figure 6.1.

This unusual protein structure was experimentally studied by Lamparter and co-workers to observe a relation between the structure and the photo-reduction of *PhrB*. Additionally, mutants of the *PhrB* wild type (WT) were studied to resolve a relation between selected amino acids and the functionality of the protein.[26,34] These mutagenesis studies[34] identified Trp342 and Trp390 as essential for CT. Therefore, the role of Tyr391, which lies between Trp390 and FAD, is unclear as its replacement by Phe, Y391F, does not block photo reduction which is reported for many mutagenesis studies of other PCF.[71,77–79] However, the mutation of Tyr391 into Alanine (Ala), Y391A, blocks the photo-reduction which underlines the relevance of Tyr/Phe at position 391 and indicates that CT occurs *via* the alternative Triad 391-390-342 of *PhrB*.

This photo-reduction ensures the activation of the protein to enable the DNA repair ability in which the negatively charged FAD will be excited and a further CT to the bond DNA damage can occur. The DNA repair activity was also studied experimentally for WT and mutants. The WT and Y391F are able to repair the damaged DNA in 5 min in the presence of Mg^{2+} ions[27,36] completely. On the contrary the Y391A showed no DNA repair activity, which was expected due to the missing photo-activated reduction of FAD. Furthermore, the mutation of Tyr391 into Trp, Y391W, was experimentally performed and showed a loss of the chromophore and FAD probably due to miss folding. Nevertheless, DNA repair assays, using an extended timescale, showed that Y391W was still able to repair 8.7% of the damaged DNA. The small percentage of DNA repair may suggest that a small fraction is still able to bind FAD, and therefore able to repair the DNA lesion *via* an ET.[26]

To clarify the time-resolved CT of the WT and mutants, classical MD and QM/MM simulations are performed. These methods allow to study the CT and the direct response of the environment with respect to the travelling charge is displayed. The molecular view of the Triad with respect to the different mutations can help to explore the pathway for the CT and can explain why the Y391A lost the CT ability.

In a second part the impact of Mg^{2+} is investigated with simulations, because experimental investigations showed a correlation with the presence of two Aspartate (Asp) residues (Asp174 and Asp254) in the binding pocket. Therefore, the two Asp residues are mutated and their impact is investigated theoretically.
This forms the starting point to deepen the understanding of the complex repair function of the 6-4-PL. First of all, the DNA and protein interaction needs to be investigated

to describe differences or similarities to previous studies based on the 6-4-PL from *Drosophila melanogaster*. Then, the repair by ET and PT can be analysed by simulations. A Histidin, His366 in *PhrB*, was reported to be essential for the DNA repair in *PhrB*, which is in a similar position than a His *Drosophila melanogaster* which is supposed to be involved in the PT and ET in other 6-4-PL. Hence, in the following studies it shall be focused on the protonation state of this His366 and the interaction with DNA is also analysed.

Taken together, the following mutations are investigated at an experimental level and therefore molecular observations will be used to explain the observations further. This will deepen the knowledge of CT processes enabling the photo-activation of *PhrB* and to further explore the molecular DNA repair machinery. Furthermore, the environment and the direct correlation with the CT paths are studied. In the following Table 6.1 the experimental mutations, which will also be simulated in the following, and their analysed impacts are listed.

Table 6.1.: Summary of the experimental investigated mutants of *PhrB*.
Lamparter and co-workes investigated the following mutations of *PhrB* and analysed their impact in the photo-activation and DNA repair activity.

Protein	photo-reduction of FAD	DNA repair activity
Wildtype	✓	✓
Y391F	✓	✓
Y391W	-	✓ (strongly reduced)
Y391A	-	-
H366N	✓	-
D174N	✓	✓ (no Mg^{2+} effect)
D254N	✓	✓ (no Mg^{2+} effect)

The experimental observations of Lamparter et al. raised the following question which will be explained in the following sections:[25,27,34,36]

- Why does *PhrB* position a Tyr at the beginning of the Triad, which has a less favourable ionisation potential than Trp?
- Is this Tyr391 residue involved in photo-reduction as the first electron donor of FAD, as proposed by its spatial position? If yes, how can the electrons be transmitted *via* this Tyr residue and why is it a Tyr residue, whereas in other groups of PL and CRY a Trp residue serves as electron donor for FAD?
- Phe is usually used to interrupt electron chains. Why does the replacement of Tyr391 by Phe not interrupt the CT path?
- How would the CT occur along a Triad, consisting out of three times Trp such as in other PCF, in this protein environment?

- Which amino acids are interacting with the 6-4-DNA lesion in the *PhrB* binding pocket?
- How does the Mg^{2+} influence the repair efficiency and how does the detected Asp residues contribute?

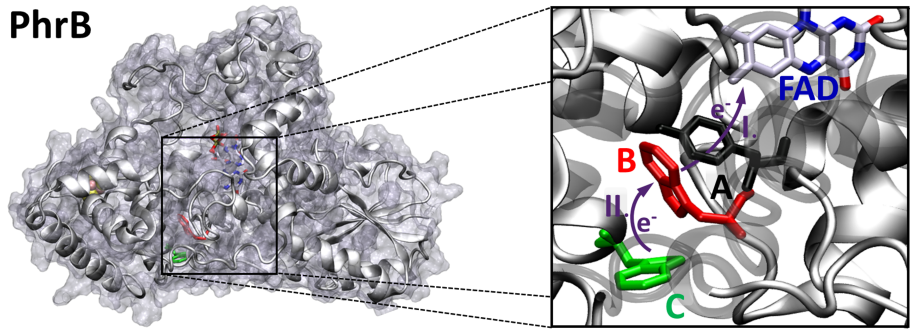

Figure 6.1.: *PhrB* structure (PDB:ID 4DJA) and charge migration pathway from the protein surface Trp342 (C, green) to FAD *via* Trp390 (B, red) and Tyr391 (A, black).[25] Light induced consecutive ETs B→FAD* and C→B are indicated by purple arrows and labeled I and II, respectively. Images were rendered with VMD.[145]

6.1. Computational details

6.1.1. Model structures and MD simulations

The structural model of PhrB WT has been derived from the X-ray crystal structure of Zhang et al. (PDB ID 4DJA).[25] To study the CT processes involving the photo-activated FAD and the Triad, several mutations were created: the *PhrB* WT structure was used as model structure and the aromatic cycle of Tyr391 was replaced by a phenyl (Y391F), or a methyl (Y391A) or an indol ring (Y391W). Two conformations of the indol ring were allowed by the protein structure, but steric hindrance prevents rotation from one to the other (see Figure 6.2). In the first conformation, termed Y391Wp (for proximal), the Trp side chain can orient toward FAD being in a closer contact than in the second conformation, termed Y391Wd (for distal).

To study the DNA repair process in *PhrB*, a bond 6-4-lesion of the crustal structure of *Drosophila* 6-4-PL[146] was placed into the binding pocked and the surrounding amino acid side chains were slightly rotated to ensure no sterical clashes. This structure was equilibrated and simulated for 1 μs to ensure no artificial created binding interaction. Furthermore, several mutations were created which affect the repair function of *PhrB*, proven by experimental studies.[34,36] Two Asp, Asp254 and Asp179, were mutated into Asn and will be called D179N and D254N.

All mentioned, classical MD simulations were performed with GROMACS 5.0.4 package[147,148] using the AMBER-SB99-ILDN force field.[149,150] The force field parameters

for neutral (oxidised) FAD and negatively charged cofactor $FAD^{\bullet-}$were taken from riboflavin and adenosine diphosphate (ADP) models developed in previous studies and taken from Ref. 16–18. The GAFF parameters[151,152] were used for the DMRL antenna chromophore and the DMRL atomic charges were calculated by restrained fitting on the electrostatic potential (RESP)[153,154] at HF/6-31G*42,43 level with Gaussian 09 package.[155] Bonded parameters of the cubic FeS-cluster were taken from Ref. 156 and the charges were taken from Ref. 157

The loop region from residues 180 to 182, which might impact the DNA binding abilities,[25] was not structurally resolved in the X-ray structure of WT *PhrB*, therefore, MODELLER[158] was used to reconstruct the loop. WT and mutated proteins (Y391F, Y391A and Y391W) were solvated in a 106.24 $Å^3$ cubic box filled by TIP3P water molecules[159] and twelve sodium ions were added to create a neutral system for the CT study. Additionally, the experimental study showed the influence of magnesium cations, Mg^{2+}, therefore, simulations with Mg^{2+} were performed for the DNA binding study.

Equilibration of the solvated proteins (WT and mutants) starts with a minimisation step, followed by 100 ps MD in the NVT ensemble and 100 ps in the NPT ensemble. 100 ns of production NPT MD simulations were performed afterwards, using the Nosé-Hoover thermostat[112] to keep a constant temperature at 300 K and Parinello-Rahman barostat[113] to keep the pressure at 1 atm. Covalent hydrogen bonds were fixed on a constant length by the use of the LINCS algorithm[160] and the time step was set to 2 fs for all classical MD simulations.

6.1.2. QM/MM simulations

In the following, the QM/MM setup is explained, which was applied to study the CT along the Triad. To treat the CT processes, a quantum mechanical treatment of the active site has to be included *via* an non-adiabatic CT-QM/MM MD, a multi-scale method, which is explained in section 5.5 in more details. The structural part of interest for CT, which was treated at QM level, contains the side chains of amino acids constructing the Triad (**A**, **B** and **C**, see Figure 6.1) and the isoalloxazine ring of FAD. The remaining atoms were treated classically using force fields (MM) and affect the QM region by electrostatic interactions. Hydrogen link atoms[161] are inserted at the QM/MM boundary, namely in the C_α-C_β bond of **A**, **B** and **C** side chains or in the C_1-C_2 bond of the FAD D-ribitol tail.

The charge propagation simulations were performed using an in-house GROMACS 4.6 version.[99] The CT between the Triad and the isoalloxazine ring was initiated by the excitation of the FAD cofactor. The first transfer continuing the excitation was studied in some PL to happen within one picosecond.[74] These two ultra-fast events were excluded in the previous studies[16–18] to reduce complexity and to focus on CT along the Trp Triad. To keep consistency, the same exclusion was also applied here and the QM region consist in **A**, **B** and **C**. Charge propagation simulations start with an electronic

state, where $FAD^{\bullet -}$ was a radical anion and the first site **A** a radical cation, $Trp^{\bullet +}$. The hole on the first site was then propagated along the Triad and moved in the reverse direction compared to the electron. These simulations were performed on *PhrB* WT and the Y391W mutants. Charge occupation of each site, varying form 0 (neutral state) to 1 (fully oxidised state) is followed during QM/MM MD simulations.
100 ns classical MD trajectories were used to randomly chose 20 (Y391W) to 25 (WT PhrB) starting structures to guarantee well equilibrated systems, while sampling different initial conditions. A kinetic model[17,18] was used to fit the average occupations (from the individual simulations) of each site and thus determine the different reaction rate constants.

6.2. Structural analysis of the charge transfer path – Triad

Classical MD, 100 ns, simulations of *PhrB* WT and the mutants (Y391F, Y391A and Y391W) are performed based on the crystal structure of *PhrB*. In these simulations, no major conformational change of the overall protein structures is observed. In the WT *PhrB* the residues Y391, site **A**, W390, site **B** and W341 site **C**, shown in Figure 6.1, form the typical structural pattern of the Triad. In the four mutants (Y391F, Y391Wp,d and Y391A) these residues occupy similar positions, therefore, the mutation does not directly involve a structural reorientation of the Triad. As shown in Table 6.2, the distances between all neighbouring sites (FAD–**A**, **A**–**B**, **B**–**C**) in WT and mutant structures are between 5 and 8 Å. These are typical nearest neighbour distances for the Triad in PCF, which were also reported in other MD simulations of CRYI of *Arabidopsis thaliana* and class-I CPD-PL of *E. coli*.[16,18] These two proteins are well-studied and will be used as comparison is the following sections, abbreviated as CRYI and PL. Thus, the orientation of the amino acids, which form the Triad pattern in *PhrB*, show the typical formation of a Triad which allows sufficiently large electronic couplings in order to enable fast CT.

Table 6.2.: Average distances between the centre of mass of the different aromatic sites (FAD isoalloxazine ring and Tyr, Phe or Trp side chain), involved in the FAD reduction in *PhrB* WT and its mutants Y391F, Y391A, Y391Wp and Y391Wd.

Distance / Å	WT	Y391F	Y391A	Y391Wp	Y391Wd
FAD–**A**	7.89	8.21	-	7.19	7.54
A–**B**	5.29	5.24	-	6.09	5.45
B–**C**	6.65	6.51	6.51	6.22	6.40
FAD–**B**	12.28	12.70	12.36	12.76	12.07
FAD–Tyr395	8.28	8.06	8.29	8.03	7.67
B–Tyr395	6.77	6.93	6.95	7.01	7.00

The experimental findings for the Y391W mutant, which has lost both chromophores, suggests a more drastic impact on the protein structure which would require simulation

protocol dedicated to protein folding and FAD docking. On the contrary, there was still a DNA repair function detected which support the fact that Y391W can also fold a functional enzyme. Therefore, the simulations allow to investigate the theoretical role of an amino acid independent on structural changes that might be imposed. Such simulations help to compare the relationship between environment and the first site in different members of the PCF. The analysis of the simulations and the experimental results can be used to explore the role of each amino acid.

The simulations for the Y391W mutant revealed two conformations of the Trp391 side chain (**A**), which is shown in Figure 6.2 and denominated as Y391Wd and Y391Wp. In Y391Wd, **A** stays continuously close to the second Trp side chain **B**, whereas in Y391Wp, **A** stays most of the time close and parallel to FAD isoalloxazine ring, but moves closer to **B** during a few nanoseconds (see Figure 6.2 and Figure 10.2 in the Appendix).

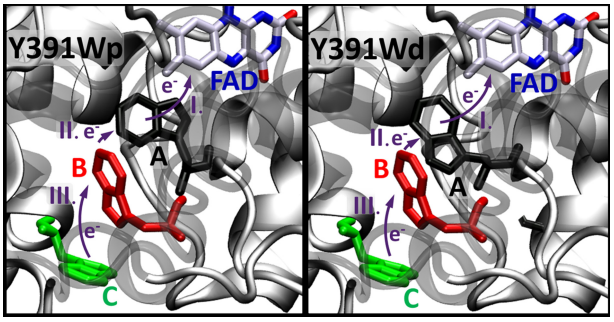

Figure 6.2.: Two conformations, proximal (p, left picture) or distal (d, right picture) of **A** in the Y391W mutant.

Environment of the first site of the Triad

The larger distances between FAD and the second neighbour between 12–13 Å (FAD–**B**) suggest that a direct ET from **B** to FAD is unlikely because the electronic coupling decreases exponentially with distance in absence of CT bridge. The distance between FAD or **B** and another Tyr, Tyr395, which is close to CT chain, ia also analysed. Mutations have no impact on the position of Tyr395, hence, it stays at around 8 Å from FAD and slightly less than 7 Å from **B** in all simulations. The analysis of the geometry of Tyr395 shows that it could influence the CT. However, Y391A does not show a photoinduced reduction of FAD_{ox} *in vitro* which shows the same orientation of Y395 than in the WT. This fact allows to discard its hypothetical influence in the CT along the Triad.

In WT, Y391F and Y391A, a water molecule which is also present in the crystallographic structure, interacts with isoalloxazine O_4 and the backbone of **A** during the whole simulation time, shown in Table 6.3 and Figure 6.4). This water, which is close to the backbone of site **A** and FAD, could form a bridge and allow a tunnelling mechanism.

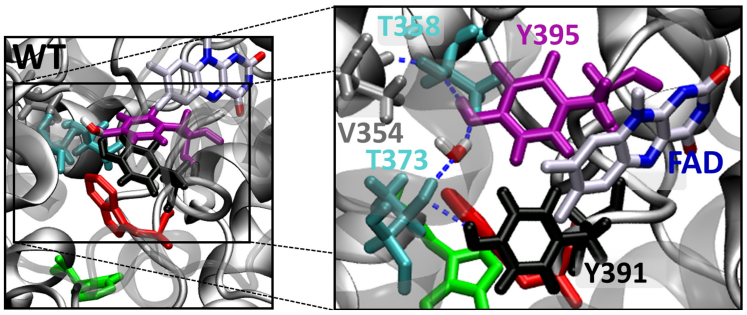

Figure 6.3.: Active site with FAD and sites **A**, **B** and **C** (left). Hydrogen bond network involving the hydroxyl groups of Tyr391, Tyr395, Thr373, Tyr395, Thr358, the backbone carbonyl group of Val354 and a structurally characterised water molecule (right)

The free space in Y391A, induced by the missing aromatic side chain of Y391, enables that more waters molecules are located between FAD and **B**. Nevertheless, one water molecule is always present in the same position as in the crystal structure and dynamically replaced five times during the hole simulation, shown in Figure 6.5. This highlights that the interaction between isoalloxazine O_4 and **A** backbone favours the stabilisation of a water molecule in this position. To estimate the effect of a water molecule bridging the FAD and the backbone of site **A** (see Figure 6.4), the pathways model from Beratan and co-workers[162] is applied. The couplings for these pathways are very small. Therefore, this pathway including the water and backbone will not be considered further (see Table 10.1 and Figure 10.3 in the Appendix).

These simulations show that the Y391A does not induced a structural reorientation of the FAD pocket directly, the space between FAD and **B** is just filled with water molecules, and one water still remains in the interacting position. In addition to the experimental observation, showing no FAD_{ox} reduction, the tunnelling *via* a bridge, consisting out of the water and backbone of site **A**, can be neglected. Thus, the aromatic ring motive is essential for the CT in *PhrB*, shown in Figure 6.5.

Table 6.3.: Average distances between the centre of mass of the crystallographic water molecule and O_4 N_5 of FAD or between the centre of mass of the crystallographic water molecule and the oxygen atom of **A** backbone, O_{bb}

Distance / Å	water$\cdots O_4$FAD	water$\cdots N_5$FAD	water$\cdots O_{bb}$
WT	2.92±0.26	3.17±0.22	3.00±0.40
Y391F	3.29±0.32	3.41±0.58	3.48±0.72

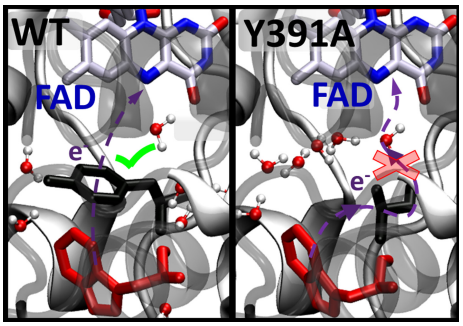

Figure 6.4.: Comparison between the CT involving site **A** and FAD in WT (left) and Y391A (right).

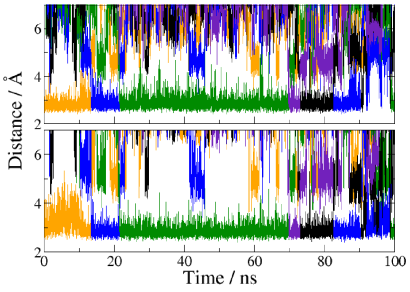

Figure 6.5.: Distance between the centre of mass of successive water molecules and FAD N_5 (top) or oxygen atom of **A** backbone (bottom).

6.3. Site energy and electronic coupling

Upon excitation of FAD by light or by energy transfer from the excited antenna chromophore, this site is reduced to the negatively charged $FAD^{\bullet -}$ species by an ET from the Triad, which then in turn becomes positively charged. In a first step, the site energies and electronic couplings are calculated to analyse the electronic structure of the neutral FAD Triad system. In a second step, the changes in electronic structure due to the CT are analysed. In previous theoretical studies of CRYI and PL, in which a similar method was used,[16–18] both proteins showed a highly exergonic and fast CT on a picosecond time scale. To characterise the CT in WT *PhrB* and to further investigate the effect of the mutation at site **A**, the relevant parameters for CT are compared with these two reference systems.

The HOMO energies of **A**, **B** and **C**, called site energies, are a direct measure of the (relative) ionisation potentials (IP) for the residue in the protein. The electronic couplings between the different partners are a measure for the CT probability, i.e. they can be directly related to the prefactors in Marcus theory.[163]
The site energies for different molecules show non-systematic errors, which can be corrected adding a constant energy shift depending on the chemical identity of the fragment. This correction was evaluated in Ref. 132 for the relative energies of Trp and Tyr and expanded in this work to Phe and FAD relative to Trp. The energy difference between the IP of Trp and Phe are calculated at DFT level and compared to the difference between the HOMO energies of DFTB. The obtained value, shown in Table 6.4, is added to the site energy to reproduce the DFT difference at DFTB level.

The CT parameters have been evaluated for oxidised FAD and a neutral Triad along the 100 ns classical MD simulations on the neutral state of the protein using the FO-DFTB/MM scheme (as discussed in the method section 5.5). Average values of energies and couplings are reported in Figure 6.6, the error bars indicate the associated

standard deviation.

According to these data, site energies of Trp390 and Trp342 (**B** and **C**) are similar, independent of the chemical nature of **A**. The electronic coupling values associated with the CT between these Trp residues are about 4.1-4.2 meV for WT and the Y391F mutant, 5.9 meV for the Y391A mutant and between 7.0 and 9.0 meV for the two rotamers of the Y391W mutant. These values are comparable with those found in CRYI and PL. The small increase of electronic coupling from WT to Y391W corresponds to a small decrease in **B**–**C** distance in the mutant (see Table 6.2). For reference, the location of the HOMO level of neutral FAD is also shown, which is the electron acceptor after excitation of one electron of the Triad.

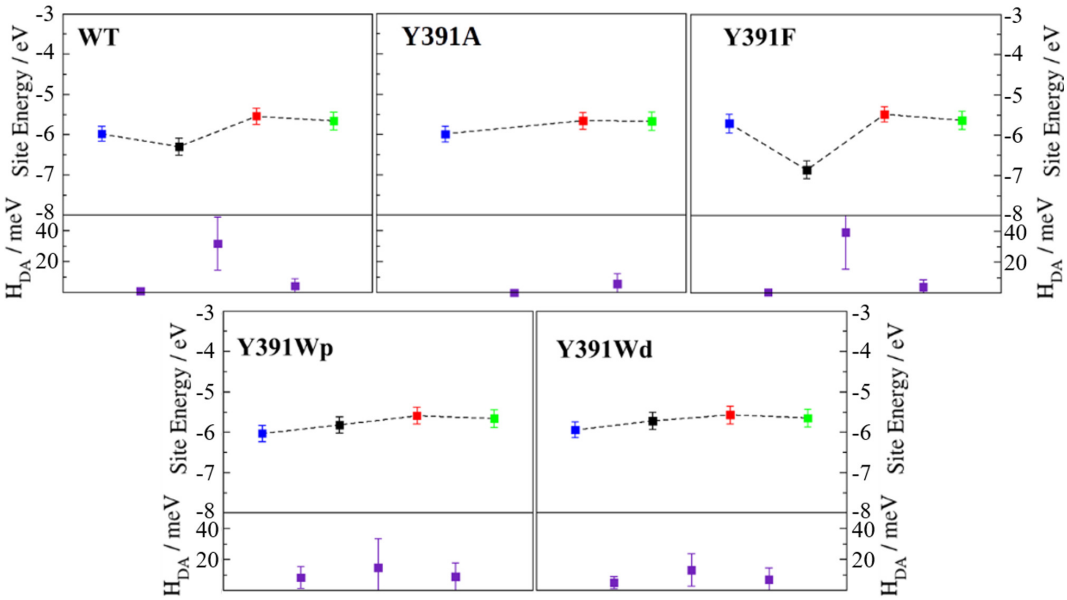

Figure 6.6.: Average site energies and average electronic couplings for *PhrB* WT and its mutants (Y391F, Y391A and Y391Wp/d). Error bars denote the corresponding standard deviation. The different symbols in each site energies panel from left to right represent values for FAD (blue) and the residues **A** (black), **B** (red) and **C** (green). Electronic coupling H_{DA} symbols (purple) are positioned between the two sites considered as donor and acceptor.

The site energy of **A** as an aromatic residue (Trp, Tyr or Phe) inside the protein follows the same order as the HOMO energies in gas phase, shown in Table 6.4.
HOMO $\approx$ -IP: $HOMO_{Trp} > HOMO_{Tyr} > HOMO_{Phe}$
In the protein, **A** has a value of about -5.8 eV for Trp, which is substantially decreased to -6.3 eV for Tyr and to around -7 eV for Phe. In the Y391W mutant, the energy of **A** is similar to energies of **B** or **C**, leading to a small energy gap between the CT partners (within about 0.3 eV). Electronic couplings are twice as large as in *PhrB* WT or Y391F. These results support the possibility of a charge migration involving a Trp Triad in the Y391Wp/d mutant. The distal conformation Y391Wd facilitates **B** $\rightarrow$ **A** CT by decreasing the **A**–**B** energy gap by 0.08 eV and the average **A**–**B** distance by 0.64 Å compared with Y391Wp (Table 6.2 and Figure 6.6). On the contrary, in the proximal conformation the

distance between the aromatic ring of **A** and isoalloxazine reduced which must enhance **A** $\rightarrow$ FAD CT.

Table 6.4.: Ionisation potential (IP) determined in gas phase at $\omega 97XD$[164]/6-31g**[154,165] level for FAD and the different side chains considered in position 391. DFTB HOMO energies are calculated for the same geometries and the shift values correspond to the difference between IP (DFT level) and HOMO (DFTB level) energy gap.

	FAD	Tyr	Trp	Phe
IP / eV	8.05	8.15	7.54	8.84
HOMO / eV	5.63	5.64	5.32	6.41
Shift / eV	0.20	0.33	0.0	0.21

As expected, the lower site energies of Tyr in *PhrB* WT and Phe in Y391F may present a significant barrier for hole transfer from FAD, as clearly seen in Figure 6.6. While previous studies of CRYI and PL showed a hopping type mechanism along the Trp Triad, in WT and Y391F a tunnelling mechanism could be expected to be in operation, which would lead to less efficient CT. Therefore, the electronic coupling between FAD and **B**, which involves **A** as a bridge, for the WT and Y391F is calculated using bridge meditated coupling method (for more details see Ref. 118 and section 5.2).
The calculation shows a coupling of 0.06 and 0.02 meV for WT and Y391F which is ten-fold more than the direct (assuming no bridge residue) coupling between FAD and **B** 0.003 meV and 0.005 meV, respectively. Electron tunnelling through the side chain of **A** seems possible and much more likely than a direct transfer between FAD and **B**, which was also indicated by the experimental studies of Lamparter on Y391F,[26,34] showing the formation of $FAD^{\bullet -}$ after photoactivation. Tunnelling involving a Phe and protein backbone has been also been described for mutants of PL,[166] therefore, in the case of Y391F, where a barrier of more than 1 eV is apparent, tunnelling has to consider definitely.

However, in WT the presence of Tyr391 causes a smaller barrier, hence, tunnelling is not necessarily the case. CT studies in DNA[167] showed that small CT barriers up to 0.4 eV may easily be overcome due to molecular fluctuations, especially for short bridges distances. The hole transfer between FAD and **A** falls in this range, allowing a direct hopping mechanism (see Figure 6.6), additionally, electron hopping *via* Tyr would involve a transiently positively charged side chain, which must be followed by deprotonation. Nevertheless, a nearby proton acceptor is missing in the *PhrB* structure, ruling out the deprotonation mechanism of Tyr391. Oxidised Tyr cannot be stabilised while the electronic coupling between **A** and **B** is strong (see Figure 6.6). If FAD $\rightarrow$ **A** transfer occurs, the following hole transfer between **A** and **B** must be very fast.

Other CT pathways can also be considered from the crystallographic structure and are also studied on an experimental level[34]. A CT chain *via* Tyr399 and Tyr40 has been suggested in an experimental study of Lamparter and co-worker.[34] Spectral changes

related to FAD reduction in Y399F mutant are slightly slower compared to WT. Nevertheless, the absence of FAD reduction in both W390F and W342F mutants clearly indicates that Trp390 and Trp342 are essential in the CT process, which cannot be compensated by a transport *via* Tyr399 and Tyr40, consequently, the Tyr399-Tyr40 CT pathway is rejected in this study.

Another possibility could be a transfer between FAD and Trp390 *via* Tyr395, visualised in Figure 6.3, which was therefore also analysed in the classical MD trajectories. This residue could substitute site **A** (Tyr391), since it has similar distances to FAD and site **B**, as shown in Table 6.2. Interestingly, these distances do not change upon mutation at site **A**, i.e. the FAD binding pocket and the connection *via* Tyr391 is very similar in all variants. According to the computed electronic couplings and site energies (see Table 10.2 in the Appendix), this pathway seems to be possible as well. However, the experimental photo-reduction studies[26] for the Y391A mutant clearly indicate the absence of any CT, i.e. the pathway does not seem to be a possible alternative (mentioned in section 6.2). Although, it has been shown that the multi-scale scheme for the calculation of CT couplings is very reliable for all couplings between the oxidisable residues.[168] The couplings between FAD and site **A** respective Tyr391 are very small, i.e. a small change in geometry could lead to a different result. For example, the *PhrB* structure in solution (experimental photo-reduction measurements) could differ from the crystal structure used for calculations. Further, the couplings are computed for neutral FAD in the ground state, whereas the active state is an excited state, FAD^{*}_{ox}, which may have some impact on the value of the coupling. As a consequence, within the accuracy of the calculations reported here, the rather small value reported for the FAD-Tyr395 coupling (see Table 10.2 in the Appendix) could as well turn out to vanish. Therefore, the conditions of the experimental study seem to impede the pathway *via* Tyr395.
However, since an analogous alternative pathway has been reported for the *PhrB* homologue CryB,[169] further investigations exploring the conditions for a possible use of this pathway in *PhrB* could be interesting. For the focus of this work, it is very interesting to note that for this alternative pathway a Tyr is the primary electron donor of FAD, i.e. highlighting the role of Tyr in the CT cascade with or without direct impact of Tyr395.

Polarisation of the environment

The focus of the following part is based on the energetics and couplings for the case, where FAD and one of the residues of the Triad is charged. The energetic landscape is changed drastically when $FAD^{\bullet-}$ and a charged $\mathbf{A}^{\bullet+}$ are considered, which is highlighted in Figure 6.7. The first ET results in negatively charged $FAD^{\bullet-}$ and a positively charged side chain on one site of the Triad and the charge separation has a sizeable effect on the site energies. Positive charge on **A**, **B** or **C** leads to a strong polarisation of the environment. This polarised environment in turn leads to a stabilisation of the

charge at the respective site. In Marcus theory, this effect is called outer-sphere reorganisation energy, λ_o (explained in section 5.3); the simulations manifest this effect by lowering the site energies with regard to the neutral states.

For each charge state on **A**, **B** and **C**, the site energies are computed along 1 ns MD simulations containing the charge on the respective site during the MD simulations, as shown in Figure 6.7. This effect was studied in more details by G. Lüdeman and B. Woiczikowski, showing that the solvent has a distinct impact, in particular to those sites which are more solvent exposed,[18] such as site **C** which is located on the protein surface. Computational studies of CRYI or PL showed that the positive charge stabilisation follows a downhill scheme, with an increasing energy gap between neutral and charge residue from **A** to **C** and where neighbouring sites had an energy difference of about 0.5 eV (Figure 6.7), enabling the fast CT in a picosecond regime.

Table 6.5.: Increase of the site energy by the environment. The charge is fixed on one site of the Triad and surrounding structures are able to adapt this new charge state for 1 ns. This shows the strong increase of the site energy. The values for WT and Y391Wp/d and for PL and CRYI (similar to the previous studies but new simulations are performed) are listed. The bold marked values correspond to the increased site energies and the last column shows the difference between the neutral and charged Triad, $\Delta_{\mathrm{Neutral-Charged}}$.

	Site Energy / eV neutral Triad		charged Triad			
			Site 1^+	Site 2^+	Site 3^+	$\Delta_{\mathrm{Neutral-Charged}}$
WT *PhrB*	Tyr391	-6.26	**-3.87**	-4.98	-5.37	2.38
	Trp390	-5.50	-4.40	**-3.66**	-4.34	1.84
	Trp342	-5.65	-5.14	-4.41	**-3.25**	2.40
Y391Wp	Trp391	-5.76	**-3.71**			2.04
Y391Wd	Trp391	-5.77	**-3.68**			2.10
CRYI	Trp400	-5.66	**-4.19**	-4.57	-5.03	1.47
	Trp377	-5.72	-4.85	**-3.77**	-4.84	1.94
	Trp324	-5.61	-4.97	-4.54	**-3.47**	2.13
PL	Trp382	-5.53	**-4.00**	-4.74	-4.89	1.53
	Trp359	-5.59	-4.46	**-3.55**	-4.36	2.05
	Trp306	-5.38	-4.78	-4.54	**-3.31**	2.07

In *PhrB*, a similar stabilisation on **C** occurs, resulting from the solvent exposure, however, surprisingly also site **A** is massively stabilised, for Tyr even more than for both Trp rotamers. In Y391Wp/d, the **A** and **B** site energies are more similar and the CT from **B** to **A** no longer follows the mentioned downhill scheme. In total, the energy difference between site **C** and **A** in Y391Wp/d is only half of the value compared to the other two

systems, which has a drastic effect on the CT equilibrium. Figure 6.6 shows the difference in site energies of all three, mentioned proteins and additionally the two Y391Wp/d mutants and Table 6.5 presents all values of the site energies due to relaxations of the environment.

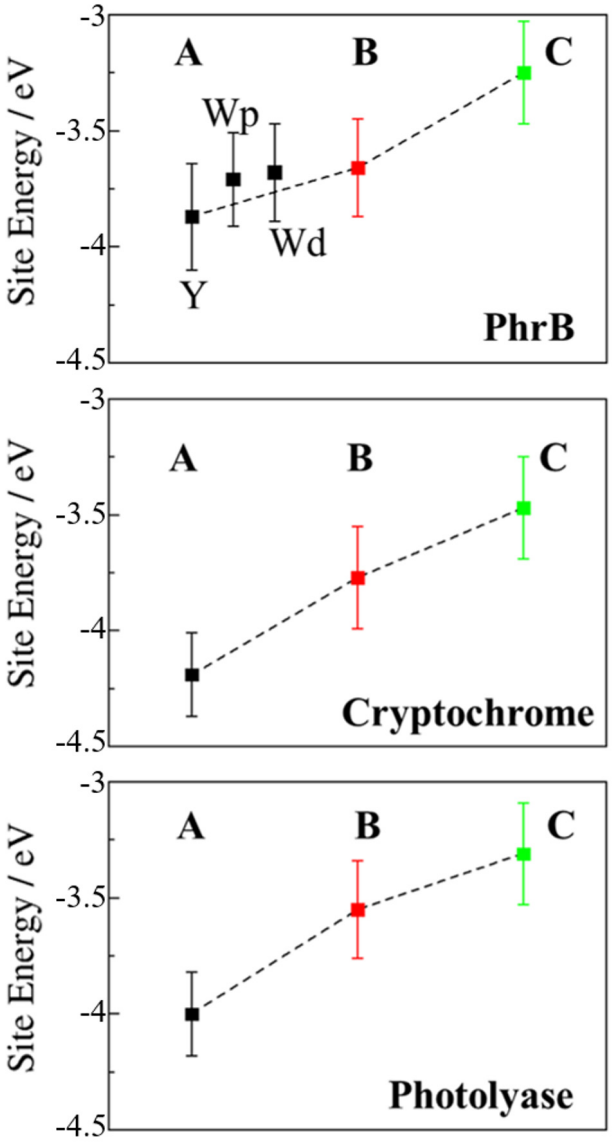

Figure 6.7.: Average site energies of positively charged **A** (black), **B** (red) and **C** (green) for PhrB, CRYI[18] and PL.[16,17] Error bars denote the corresponding standard deviation. For PhrB site **A**, the energies labelled with Y, Wp and Wd correspond to WT, Y391Wp and Y391Wd respectively.

To compare the same proteins, which were also studied in Ref. 17,18, the simulations are performed again to ensure that the same simulation protocol is used and the same results are carried out. An obvious energetic difference for the site **A** Trp in Y391Wp/d, PL and CRYI is observed. In Y391Wp/d, **A** is nearly isoenergetic to **B**, while in the other proteins, positively charged **B** is more stable than **A** by about 0.4 eV. In Ref. 17,18

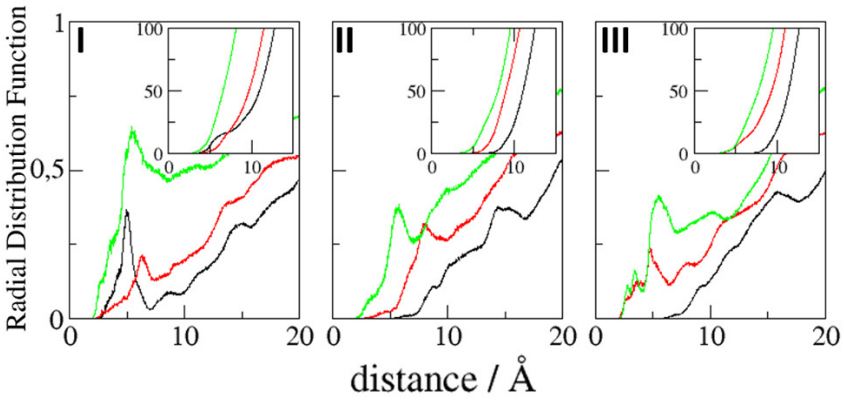

Figure 6.8.: Radial distribution function of water molecules around **A** (black), **B** (red) and **C** (green) in *PhrB* WT shown in **I**. , PL in **II**., CRYI in **III.**. Integration of the radial distribution curves is given in insets, presenting the number of water molecules around the sites with respect to the centre of mass distance.

structural reasons for this energy gap between **A** and **B** are analysed. In PL or CRYI, **A** is buried in a pocket with more than 5 Å distance from any water molecule, as documented by calculation of water distribution functions along MD simulations, visualised in Figure 6.8. On the contrary, water molecules can easily move toward **A** in *PhrB*: the radial distribution function of water around Tyr391 presents and a peak around 5 Å, while the first peak around **B** is observed at 7 Å (Figure 6.8). Moreover, the classical MD simulations carry a stable hydrogen bond network formation which can also participate in $\mathbf{A}^{\bullet+}$ stabilisation. Figure 6.3 shows the construction of the hydrogen network which contains Thr373, a water molecule, Tyr385, Thr358 and the backbone of Val354.

In PL and CRYI, sites **B** and **C** are stabilised by solvent interactions, leading to a larger solvent driven reorganisation energy. This explains the downhill energetics as shown in Figure 6.7. On the contrary, water molecules close to the FAD-**A** complex in *PhrB* help to stabilise the charge-separated RP-**A** ($\text{FAD}^{\bullet-}$ $\mathbf{A}^{\bullet+}$) state, and compensate the unfavourable intrinsic IP of Tyr. Likewise, this could impact the CT efficiency in the Y391Wp/d mutants by increasing the probability to localise the charge on the first member of the Triad.

6.4. Direct charge transfer QM/MM simulations

To study the implication of the energetic landscape on the CT dynamics, unbiased simulations of the CT through residues **A**, **B** and **C** in WT protein and Y391Wp/d rotamers are performed. Y391F and Y391A are not considered in this part as Phe and Ala cannot be oxidised. These simulations indicate the formation and lifetime of each radical pair state: RP-**A** ($\text{FAD}^{\bullet-}$ - $\mathbf{A}^{\bullet+}$), RP-**B** ($\text{FAD}^{\bullet-}$ - $\mathbf{B}^{\bullet+}$), RP-**C**($\text{FAD}^{\bullet-}$ - $\mathbf{C}^{\bullet+}$), summarised in Figure 6.9. The averages over 25 simulations for WT and 20 simulations for the

two Y391Wp/d mutants is discussed in the following are presented in the left part of Figure 6.9 and the summary of all individual simulations is shown in the right part of Figure 6.9, in which the different colours correspond to one individual simulation and represents therefore the percentaged occupation of the individual sites.

The charge is considered to be on a specific site when the occupation of the site is larger than 50%. The time dependence of the average site occupations have been fitted using a kinetic model, which was also established in Ref. 16–18, to obtaining rate constants, which are listed in Table 6.6:

$$A \underset{k_{21}}{\overset{k_{12}}{\rightleftarrows}} B \underset{k_{32}}{\overset{k_{23}}{\rightleftarrows}} C \tag{6.1}$$

Therefore, the two-step kinetic process can be described by the following coupled rate equations, including forward and backward transfer:

$$\frac{d[A]}{dt} = -k_{12}[A] + k_{21}[B] \tag{6.2}$$

$$\frac{d[B]}{dt} = -k_{23}[B] - k_{21}[B] + k_{12}[A] + k_{32}[C] \tag{6.3}$$

$$\frac{d[C]}{dt} = -k_{32}[C] + k_{23}[B] \tag{6.4}$$

The brackets [..] denote the normalised concentration of the sites, which can be also seen as the population of the site. The starting conditions for the CT assume an initial population of just site **A** which corresponds to:

$$[A]_{t=0} = 1 \quad [B]_{t=0} = 0 \quad [C]_{t=0} = 0 \tag{6.5}$$

Then, Eq. 6.3,6.4 and 6.4 are solved numerically adopting the rate constants until the deviation of the used kinetic model fits to the averaged populations.

In WT, the population of the first radical state RP-**A** drops within a few ps and no back transfer is observed. The second state, RP-**B**, is transiently occupied by 60% of the total positive charge before decay and formation of the third state RP-**C**. After 25 ps simulation time, the charge distribution remains stable, with roughly 85% of the charge on the third state, while 15% remains on the second state.
Due to the averaging over several simulations, these numbers show a statistical distribution rather than a charge delocalisation over different sites. The positive charge is therefore well stabilised on the solvent accessible Trp342. No backward CT from **B** to **A** occurs during 1 ns of simulation time. The very fast first CT from Tyr391 to Trp390 shows a significantly larger transfer rate than that calculated for plant CRYI. In *PhrB*, k_{23} and k_{32} are consistent with values from CRYI and PL, with a backward transfer 10-fold smaller than the forward one.

The Y391Wp/d mutants show a different behaviour. On average, the hole remains on

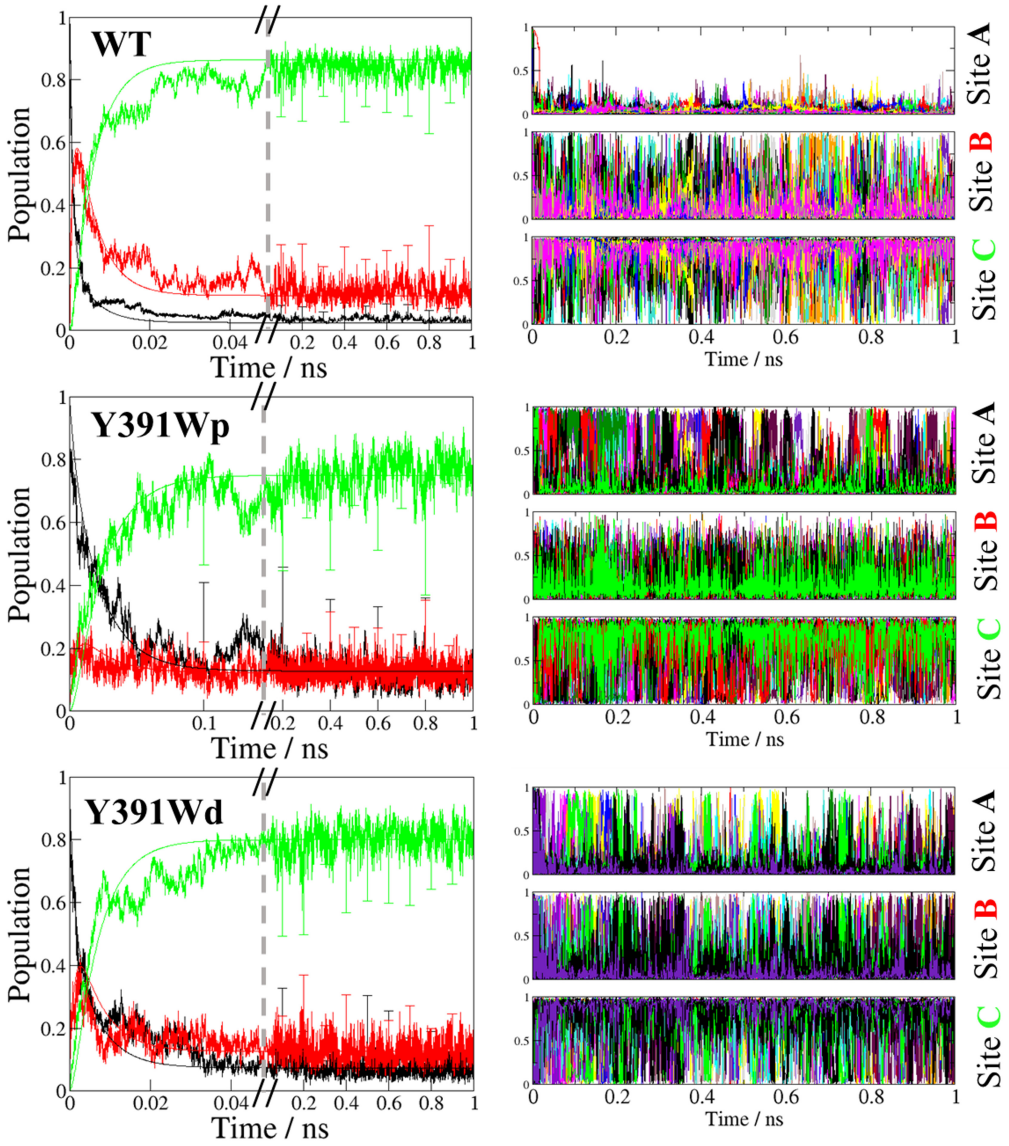

Figure 6.9.: Time dependent evolution of the charge occupation along the Triad in *PhrB* WT, Y391Wp mutant and Y391Wd mutant. Black line: RP-**A**; red line, RP-**B**: green line: RP-**C**. The left graphs show the average occupation from all simulations of 1 ns length for WT or Y391Wp/d mutants, respectively. Error bars denote the standard deviation each 0.1 ns. The curves resulting from a fit using a kinetic model, shown in Eq. 6.1 [18] as described in the text are also shown using the same colours. The right, stacked graphs show the population of 25 or 20 individual simulations and each colour presents an individual simulation, furthermore, the first graph corresponds to site **A**, the second to site **B** and the third to site **C** (similar for each of these stacks)

RP-**A** during the first 50 ps in Y391Wp and during the first 21 ps in Y391Wd, respectively. The final stabilisation of 70%- 80% of the charge on site **C** occurs after 70 ps for Y391Wp and 33 ps for Y391Wd. For the **A**-**B** transfer, the forward and backward rate constants are very similar in each rotamer simulation, as shown in Table 6.6. In Y391Wd, the k_{21} rate is also close to k_{23}. For the transfer from **B** to **C**, the forward rate is 10-fold higher than the backward rate. All rate constants are in the same order of magnitude as those of CRYI and PL. The main difference between the Y391Wp/d mutants and CRYI is the strong back transfer rate from **B** to **A** (see k_{21} in Table 6.6).

Table 6.6.: Hole transfer rate constants described in Eq. 6.1 for *PhrB* WT, Y391Wp/d mutant rotamers, PL and CRYI. A polarization parameter of 1.4 is used.[16]

ns^{-1}	*PhrB* WT	Y391Wp	Y391Wd	PL[17]	CRYI[18]
k_{12}	870	63	237	88	190
k_{21}	193	66	256	87	85
k_{23}	199	124	231	64	70
k_{32}	25	20	36	6	10

Table 6.7.: Average number of backward transfer from **B** to **A**, ν, and average total time τ of occupied site **A**. The charge is considered to be on site **A** when the occupation of the site is larger than 50%. A transfer is counted in τ when the charge stays more than 500 fs on site **B**. 20 CT simulations of 1 ns are taken into account for Y391Wp,d and 22 for PL, respectively.

	Y391Wp	Y391Wd	PL[17]
ν	29.8	35.2	3.5
τ(ps)	92.0	22.5	67.5

Indeed, during all CT simulations of different CRYI and PL proteins, several backward transfers to **A** are observed, however, no backward transfer is present in *PhrB* WT simulations, as the first residue is a Tyr. Thus, the number and the stability of backward transfers, shown in Table 6.7, is calculated for different systems: Y391Wp/d and PL.

In Y391W, 30 crossings in which the positive charge moved back to **A** and formed a stable RP-**A** for at least 500 fs are observed, which is ten-fold more than in PL. The difference between the two conformations of Y391Wp/d is related to the **A**–**B** distance: in Y391Wd, **A** is closer to **B**, which facilitates backward and forward CT between them. On the contrary, in Y391Wp, **A** is closer to negatively charged FAD which contributes to stabilise the positive charge on **A**. Furthermore, a charge recombination on the isoalloxazine ring becomes more likely. Nevertheless, in 100 ns MD simulation of Y391Wp, motion of **A** to distal conformation is observed and can also contribute to an enhanced CT between the two Trp residues. In both conformations, the charge is more often back transferred to **A**, but stays less time for one transfer than in PL. Therefore, the simulated trajectories and data published in Ref. 17 are analysed and taken as reference.

6.5. DNA binding and repair of *PhrB*

The previously described photo-activation of the FAD ensures that the negative charge can be placed on the FAD. This allows a further CT to repair the bond, damaged DNA. Thus, the study of the binding and repair of this 6-4 photo-adduct is the logical next part to focus on. Due to the complex mechanism of 6-4-PL compared to the CPD-PL the nature of the binding pocket needs to be analysed, which is the focus of this section. Thus, the crystal structure of *PhrB* (PDB:4DJA) is combined with a 12 nucleotides long DNA double strand having a 6-4 photo-adduct, taken out of a crystal structure of *Drosophila melanogaster* 6-4-PL.[146] Smaller rotations and adjustments of several amino acid side chains in the binding pocket are performed to overcome sterical clashes. Afterwards, free MD simulations of 1 μs are performed in collaboration with Dr. N. Gillet, to enable relaxation and equilibrate a natural and not forced binding position. The DNA PL complex stays stable along this simulation.

Mutation studies were focused on the dependence of the DNA repair efficiency with respect to selected amino acids:[34] Tyr430 and Tyr424 showed an impact in the repair and the binding affinity of DNA. Histidine, His366, is essential for the repair, similarly to eukaryotic 6-4-PL which also have conserved His. In many studies consequently the influence and interplay of the conserved His and potential water wires and the required PT and ET reactions are ruled out.[170–175]

6.5.1. Influence of MG^{2+} and the protonation state of His366 in WT *PhrB* and mutants

The main goal is to resolve the explicit impact of His366 in the repair 6-4 photo-adduct, however, the general setup needs to be analysed previously. Thus, two states of His366 are considered (positively charged $His366^{+}$ and neutral) to study the binding pocket at a molecular level and to clarify the setup for later simulations of the PT and ET processes.

Additionally, experimental studies by Lamparter and co-workers showed a dependency between the DNA repair and the presence of MG^{2+}, just affecting members of the FeS-BCP group (CryB and *PhrB*)[36] and not PL or CRY of other subgroups. This repair dependency was further investigated by mutations affecting the DNA binding pocket. Amino acids, which are highly conserved in FeS-BCP and carry a negative charge, were focused such as Asp254 and Asp179 and mutated into Asn, called D254N, D179N.[27] Simulations of D254N and D179N are performed to study their impact, additionally, the conserved His366 plays a central role and especially the protonation state might be important. Therefore, both mutants are created once with protonated His366, $His366^{+}$ and additionally with a neutral His366, which has the proton at the N_δ nitrogen, His366Nδ. The simulation having a positively charged $His366^{+}$ show the following: two of the completely randomly inserted Mg^{2+} ions enter the DNA binding pocket close to the 6-

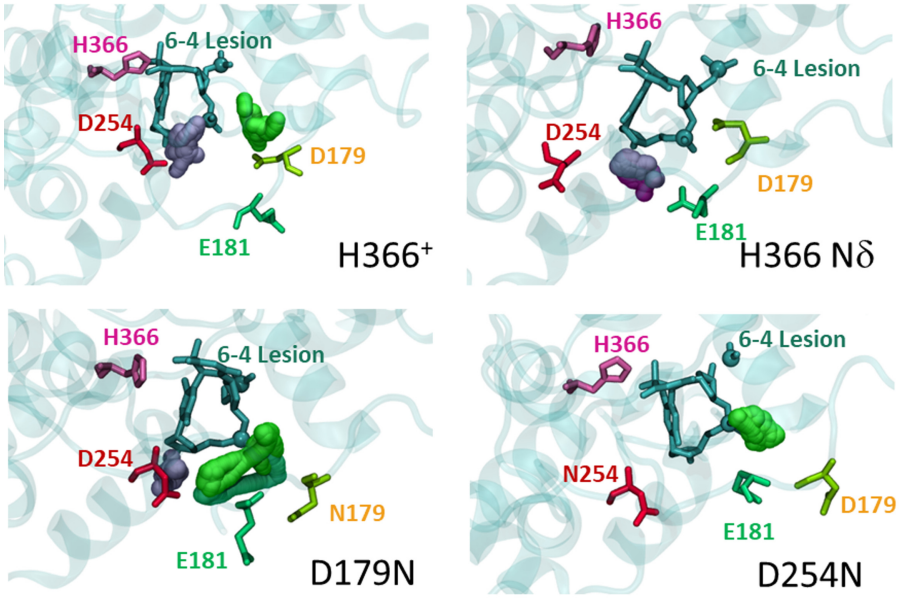

Figure 6.10.: Positions occupied each nanosecond by MG^{2+} cations along the 1 μs simulations in WT *PhrB* with positively charged or neutral His366, D179N or D254N mutants are illustrated. The cations in position Mg_a^{2+} and Mg_b^{2+} are represented by green and purple spheres, respectively. Amino acid residues and protein backbone are taken from a single snapshot along the simulation.

4 photo-adduct after a few nanoseconds and interact with the deprotonated Asp179 and Asp254. The positions which are occupied by the Mg^{2+} cations are marked by spheres in Figure 6.10 for four different MD simulations (His366$^+$,His366Nδ, D254N and D179N). One Mg^{2+} cation, called Mg_a^{2+}, engages with the phosphate group of the 3' thymine and interacts with Asp179 and Glu181. A second one, termed Mg_b^{2+}, is deeper buried in the active site and interacts with the phosphate of the 3' thymine, the oxygen atom of the 3' thymine ring and with Asp254 (see Table 10.3 in the Appendix, which presents the Coulomb interaction energies).

This allows to mention that WT *PhrB* with His366$^+$ offers two binding positions for a divalent cation, Mg^{2+}, enclosed to the 6-4 photo-adduct. On the contrary, the MD simulation, using neutral His366 in WT *PhrB*, shows a slightly different conformation of this side chain as compared to mentioned protonated His366$^+$, and the binding of only one Mg^{2+} in the active site, close to the 3' thymine, visualised in Figure 6.10. The cation, however, mainly interacts with Glu181, whereas the electrostatic interactions with Asp179 or Asp254 are very small and underline the absence of strong ionic interaction between Mg^{2+} and these two aspartates. Actually, Asp254 interacts with Arg187 during the first 750 ns of the simulation while Asp179 presents strong hydrogen bonds with Arg183. In the D254N mutant, the position at Mg_b^{2+} is no longer occupied while the Mg_a^{2+} is still presented and interacts even stronger with Asp179 or Glu181 than in WT *PhrB*. The D179N mutant presents an unexpected behaviour: the Mg_a^{2+} position is occupied all along the 1 μs MD simulation, partially stabilised by an interaction with

Glu181, while Mg_b^{2+}, presented at the beginning, is released after 500 ns. The simulation, having D179 or D254 mutated, show that the stability of the binding positions of Mg^{2+} cations close to the photo-adduct are directly affected by the lack of the negative charges at the aspartates. Additionally, the neutral His also just allows the stabilisation of one cation permanently. Thus, the simulation results are in agreement with the experimental observations, in which the D254N and D179N mutants showed a loss of the Mg^{2+} DNA repair effect. Additionally, it can be assumed that the protonated His is the most likely state for an efficient DNA repair.[27]
Therefore, the simulation is also performed in similar fashion but containing 24 sodium ions instead, which results in larger distances similar to the neutral His366 (see Figure IV in the Appendix). Due to the fact that ET from $FADH^{\bullet}$ to the damage triggers the catalytic cycle and as the ET rates strongly depend on the donor-acceptor distance, the presence of the divalent cation, Mg^{2+}, in direct interaction with the damaged thymines increases the EA of the 6-4-lesion by 280 meV, according to DFT calculations (using a LC-functional$\omega 97XD$[164] and 6-31G** basis set[154,165]). The 6-4 photo-adduct and FAD are included and the Mg^{2+} presented as point charges are used to estimate the difference of the EA with presence of the cation.

6.5.2. Structural comparison of the DNA binding pocket between 6-4-PL from *Drosophila melanogaster* and *PhrB*

Structural comparison with the eukaryotic (6-4) PL from *Drosophila melanogaster* shows that the two Asp residues (Asp179, 254) are in positions, enclosed to the positively charged to Arg421 and Lys246, respectively, shown in Figure 6.11. It was reported that these side chains have a key role in binding of lesion DNA.[146,176] Additionally, in a 6-4-PL of *Xenopus*, the homologous Arg is shown to be involved in binding the DNA lesion.[177] At the position of Arg421, the homologous animal CRYs have a His, which can be neutral or positively charged at physiological pH.[177] The positive charge at Lys246 spatial position is highly conserved in eukaryotic 6-4-PL. According to WT *PhrB* simulations, Mg_a^{2+} interacts with phosphate groups of the DNA in the periphery of the active site and should help to maintain the lesion in an optimised conformation to facilitate its flipped-out conformation. Indeed, in the presence of Mg_a^{2+}, the 6-4 photo-adduct is closer to both aromatic cycles of the FAD cofactor.

In the crystal structure of the *Drosophila melanogaster* 6-4-PL complex with lesion DNA, the Lys246 side chain is orientated to form hydrogen bonds with the 5' ring, whereas Mg_b^{2+} in the simulations interacts with the 3' ring. Nevertheless, MD simulations of *Drosophila* 6-4-PL showed a motion of the lysine toward the 3' phosphate group and thus closer to the 3' thymine ring,[178] which is in better agreement with the presented Mg^{2+} binding conformation. D179N and D254N mutations affect the Mg_b^{2+} complex in the simulations, which is consistent with a weaker stabilisation of the negatively charged

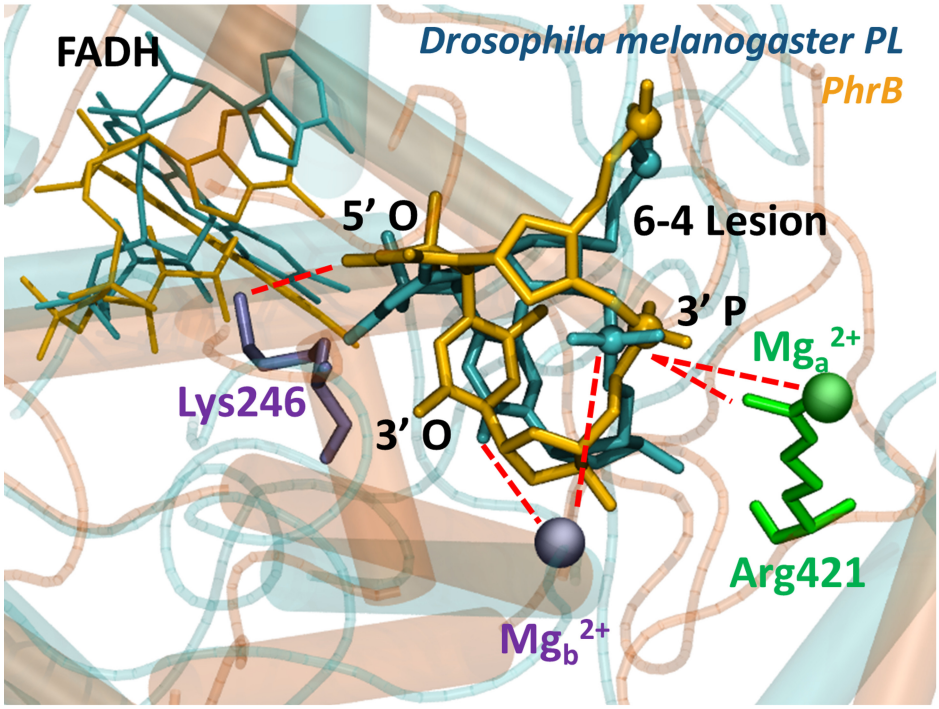

Figure 6.11.: Comparison between *PhrB* (cyan, MD snapshot) and *Drosophila melanogaster* PL (orange PDB:3CVU)[146] active site containing the 6-4 photoproduct (shown as sticks, phosphorous atoms are represented by balls), FAD (smaller sticks), and positively charged groups: Arg421 (from *Drosophila melanogaster* PL) and equivalent Mg_a^{2+} (from *PhrB*) in green; Lys246 (from *Drosophila melanogaster* PL and equivalent Mg_b^{2+} (from *PhrB*) in violet. The possible electrostatic or hydrogen bonding interactions are represented by dashed red lines. It has been shown previously by MD simulations that Lys246 can interact with the 3' phosphate group.[176]

6-4 photo-product intermediate during the repair catalytic cycle than in WT *PhrB*. The D179N and D254N mutant results support the hypothesis that Mg_a^{2+} acts close to the DNA binding site and assuming the charged His366 two $Mg_{a,b}^{2+}$ are bound to the relevant region of the protein. A direct interaction of protein bound Mg^{2+} with a phosphate of the lesion DNA is also evident. For both CPD and 6-4-PL the repair is initiated by electron flow from $FADH^-$ to the lesion. In case of 6-4 photo-product the electron is initially located on the 5' thymine at position 3, 4 or 5.[179,180] The distances between the Mg_a^{2+} cations and the isoalloxazine centre or the adenine centre of FAD are 15-17 Å, whereas the distances between Mg^{2+} and the position 3, 4 or 5 of the 5' lesion thymine are shorter, i.e. 8-9 Å which is shown in Figure 6.11.

Therefore, the positive charges of $Mg_{a,b}^{2+}$ direct the electron of $FADH^{-*}$ towards the DNA lesion and increase the electron affinity (EA) of the lesion thymine. Without Mg^{2+} the lesion is less appealing for the excited electrons from $FADH^{-*}$ and inefficient back ET becomes more likely. The ET is most likely followed by a rapid PT from $His366^+$ to the 6-4 photo-adduct. According, the change from positive to neutral His366 results in the loss of a Mg^{2+} in the surrounding of the DNA-lesion binding site. The loss of two positive charges will facilitate electron back transfer from the lesion to $FADH^\bullet$.

In summary, the Mg^{2+} system could have a beneficial effect on the DNA repair by *PhrB*

in three ways, (I) coordination of the lesion within the protein pocket and DNA binding, (II) charge stabilisation after ET from $FADH^-$ to the DNA-lesion and (III) supporting the backward ET from the 6-4 photo-adduct to $FADH^\bullet$ upon deprotonation of His366.

6.6. Conclusion & Discussion

In most members of the PCF, the central ET pathway contains a Triad of Trp residues, from the surface of the protein to the FAD chromophore. The mechanism commonly accepted for the FAD photo-reduction consists of three successive CT steps, called photo-activation and allows the proteins to occupy their active form: $FAD^*_{ox} \rightarrow \mathbf{A}$, $\mathbf{A} \rightarrow \mathbf{B}$ and $\mathbf{B} \rightarrow \mathbf{C}$.[181] Like other members of the PCF, *PhrB* is able to reduce FAD upon light absorption *via* a long range ET, involving aromatic residues. Site directed mutagenesis experiments have shown that Trp390 and Trp342, are essential for the reduction process. The distance between the isoalloxazine ring and Trp390 is roughly 12-13 Å and thus too far from FAD for direct CT. Furthermore, no other Trp in the surrounding structure can complete the Triad, rising the question of the role of the closest residue to FAD, which can be involved in CT. A Tyr, Tyr391, is situated between FAD and Trp390 in a suitable place to participate in the CT, but its oxidation appears not required for FAD reduction, because mutation of Tyr391 to redox inert Phe residue neither blocks FAD photo-reduction nor DNA repair. However, in 464 *PhrB* homologues, this residue is either Tyr or Phe, which is visualised in Figure 10.1 in the Appendix, underlining a systematic presence and probably linked functionality.

Classical MD simulations of the WT and the mutants (Y391, Y391F, or Y391Wp/d) show that the aromatic cycle is parallel to the isoalloxazine ring of FAD, in a suitable conformation for $\pi - \pi$ orbitals interaction. A CT *via* Tyr is usually not favourable due to the higher IP, which is 0.6 eV larger than for Trp (see Table 6.4). Additionally, the oxidised state of $Tyr^{\bullet+}$ requires a stabilising PT normally. However, no proton acceptor is in the neighbourhood of Tyr391, hence, an efficient hopping mechanism, where Tyr391 is transiently oxidised, seems improbable at first sight. The oxidised state of Tyr391 gets stabilised by the protein environment and a hydrogen-bond network, displayed in Figure 6.3, can attract the proton and allow the O-H bond elongation to compensate an electronic density decrease on the cycle. Thus, energy of the Tyr391 HOMO gets significantly reduced in *PhrB*. Therefore, CT from FAD to **B** *via* a transiently oxidised Tyr seems to be possible, in addition to less efficient tunnelling. Nevertheless, a much more efficient hopping regime, which is present in other PCF members, seems to be enabled in *PhrB* due to the specific protein environment of site **A**.
Mutation of Tyr391 to Phe disables this direct transfer pathway, but does not impede a tunnelling mechanism. The performed calculations show just a small direct electronic coupling between FAD and **B** and a large energy gap between FAD and Phe391. Therefore, a superexchange tunnelling mechanism through Phe is a viable hypothesis

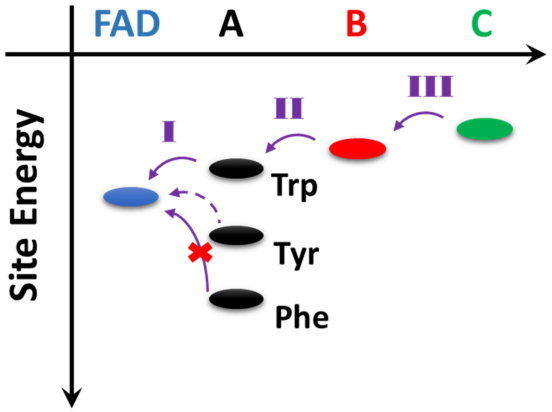

Figure 6.12.: Relative site energies: from left to right the relative positions of the site energies of FAD, **A**, **B** and **C** are visualised. For site **A**, three residue types have been investigated, Trp, Tyr and Phe.

to explain the experimentally observed CT between FAD and Trp390 in Y391F mutant.[14] Including the aromatic cycle of Phe as bridge into the coupling calculation results to a ten-fold increase of the coupling, compared to a not bridged direct FAD-**B** CT.

In general, mutation of Tyr or Trp to Phe is often used in order to impede the CT and identify redox active residues. However, experimental mutation of Tyr391 to Ala blocks FAD photo-reduction and DNA repair, underlining the relevance of an aromatic side chain in this position for the CT process. Consequently, mutation of a member of the Triad to Phe does not necessarily prevent CT even if it can impact the CT rate and modifies the CT mechanism significantly. Furthermore, this observation rises the question about the presence of alternative pathways e.g. through Tyr395 or water molecules. The MD simulation show that in Y391A the water distribution between site **B** and the backbone of Ala391 stays similar, compared to the WT. This pathway would still exist in Y391A, but the experimental study did not show a CT and therefore the alternative pathway including water and the backbone can be neglected. The analysis of Tyr395 shows in principle that it could also participate with weaker couplings than Y391. However, the same experimental study combined with the MD simulations can be used to neglect this pathway. In the Y391A mutation, no observed effect targets Tyr395, which is visible in the same obtained electronic coupling, the same the H-bond network and a similar water distribution.

The molecular insight reveals that the mutation does not directly affect the geometry around Tyr395. The CT should not be hindered, and therefore, the missing photo-reduction of Y391A (experimentally examined) and visualised molecular insight allows discarding a possible CT in which Tyr395 is directly involve. Most PL and CRY carry a Trp Triad offering a guided CT to the protein surface: in these cases, the strong exothermicity of the CT results from stabilisation of the positive charge by the solvent.[16–18] Hence, the question arises, why *PhrB* makes use of a Tyr at site **A** instead of a Trp

residue. ET can be described by an energy landscape, as shown schematically in Figure 6.12. It is clear that a Tyr substitution introduces significant barriers into the CT pathway when it is placed into the photo-reduction pathway of e.g. *E. coli* PL. In *PhrB*, however, **A** (Tyr or Trp) shows a stronger stabilisation due to interactions with water and the protein environment.
Compared to other members of the PCF, which have already been simulated,[16–18] site **A** in *PhrB* is very close to water molecules. This polar environment compensates the intrinsic IP of Tyr and results in a downhill CT from FAD to **C**. On the contrary, the Y391Wp/d mutations, based on the lower intrinsic IP of Trp, would disrupt the downhill energetics and allow a more likely charge recombination of FAD.

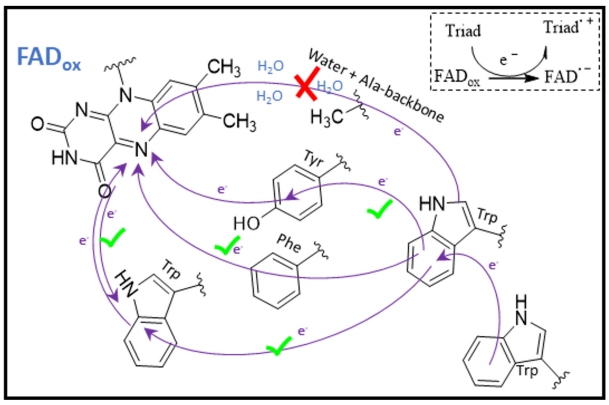

Figure 6.13.: Summary of the different investigated molecules for the first position of the Triad in *PhrB*: the purple arrows indicate the direction of the ET, the green check marks a successful transfer and the red cross a inhibited transfer path.

Therefore, a theoretical investigation, modelling the Trp Triad in *PhrB*, provides valuable insight of the importance of the Tyr residue in comparison with previously studied CRY and PL. The radical pair state of RP-**A** is stabilised in the Y391Wp/d simulations due to strong electronic couplings and small energy gaps between Trp390 and Trp391 and charge is transferred to Trp342 within 33-70 ps, see Figure 6.9. Although rate constants for the hopping mechanisms in the Y391W mutant of *PhrB* are comparable with the Trp Triads in PL and CRY[17,18] backward CT to *A* occurs more frequently in the *PhrB* mutant than in these two proteins. Such back transfer increases the risk for charge recombination on FAD and hence a more likely inefficient CT mechanism. If a Trp residue corresponded to the first site the environment would stabilise the RP-**A** state in *PhrB* obviously more than in Trp Triads of other members of the PCF.

The simulations with the included DNA lesion highlight that the conserved Asp (Asp179, Asp254) stabilise the divalent Mg^{2+} cations enclosed to the 6-4 photo-adduct. This is not required for other studied 6-4-PL because they already show positively charged amino acid side chains in similar positions, highlighted in Figure 6.11. This explains the

correlation between the repair efficiency and the cation Mg^{2+} concentration. Thus, the positive charge effectively directs the electron of $FADH^{-*}$ towards the bond 6-4 photo-adduct, resulting in an increased EA of the lesion thymine. Without the positive charge at the position occupied by Mg^{2+} the forward CT from $FADH^{-*}$ to the DNA-lesion is less supported and especially less stabilised, therefore, an unproductive backward CT becomes more likely. Furthermore, the CT is most followed by a rapid PT from $His366^{+}$ to the 6-4-lesion DNA. This determines the conditions which offer the chance for an efficient DNA repair in *PhrB* and can be used to study the ET and PT or a possible coupled PT/ET.

Taken together, the theoretical results can be used to further describe the CT in *PhrB*: the protein environment of *PhrB* is quite different compared to other members of PCF, especially around the Triad and the DNA binding pocket. Residues bigger than Tyr at position **A** result in loss of FAD binding site of *PhrB* (experimentally). However, the simulation show a stable pocked by a direct insertion of the Trp residue, pointing out a failure in the folding but not in the stability of a folded protein. Solvent can come closer to **A**, stabilising the $FAD^{\bullet -}\mathbf{A}^{\bullet +}$ state due to a large reorganisation energy. The presence of a Tyr residue instead of a Trp at this first site preserves the structure, the energetics, and therefore the functionality of the photo-activation process in *PhrB*. Due to the combination of the experimental observations and theoretical calculations, the CT processes of the ancestral prokaryotic 6-4-PL and their mutants are resolved and summarised in Figure 6.13, highlighting that the aromatic side chain in the first place of the Triad is essential for a photo-activation, involving a CT along the conserved CT path.

7. Charge transfer simulation in a class-III CPD-Photolyase – *PhrA*

Reproduced in part from Ref. 182

Biological Relevance of Charge Transfer Branching Pathways in Photolyases

with the permission of the Chem. Sci. Owner Societies.

(submitted)

In the class-I CPD-Photolyase of *E. coli*[59] the first Triad, consisting out of three tryptophan (Trp) residues was identified to enable a charge transfer (CT) from the photo-activated fully oxidised FAD, FAD_{ox}, to the protein surface. Many recent experimental and computational studies addressed this photo-reduction mechanism in several Cryptochromes (CRY) and Photolyases (PL).[16–18,26,52,64,77,181,183–187]
This CT is an essential subnanosecond process and allows the enzyme to assume its active form *in vitro*. Since the first description of the CT along the common Trp Triad in PL, different additional structures were reported, offering further possible CT pathways: a fourth Trp[181,183] can extend the Triad and Tyrosine (Tyr) can also be included and participate in the photo-reduction mechanism along the Triad.[26,77] Thus, the Triad can be elongated or an alternative branching can cause a CT into another part of the protein. The understanding of these different transfer patterns, which influence the CT cascade, can help to clarify the evolution within this protein family.

In the previous Chapter 6, the main topic was the characterisation of the CT along the Triad in a PL, called *PhrB*, from *Agrobacterium tumefaciens* and mutants of this protein. Additionally, another PL, called *PhrA*, was found in this organism. *PhrB* a 6-4-PL and *PhrA* a class-III CPD-PL, belong to the different classes of the PL enzymes. *PhrA* and *PhrB* were identified to form together an essential DNA repair system because the combination allows to repair two different sun induced dimerised DNA damages in the *Agrobacterium tumefaciens*.[24] *PhrA* has at least two possible CT pathways, participating in the FAD photo-reduction, which have been characterised experimentally and are

visualised in Figure 7.1. The first pathways is formed out of Trp384, Trp361 and Trp308 (respectively **A**, **B1** and **C1** shown in Figure 7.1), which corresponds to the classical Trp Triad and is extensively studied in many different members of the PCF. The alternative Triad includes Trp384, Trp318 and Trp367 (**A**, **B2** and **C2** also shown in Figure 7.1). In addition to that, a fourth Trp, Trp336 mentioned as **D2**, is identified close to Trp367 and was not considered in the previous experimental studies of Lamparter et al..
This Trp will be included to this theoretical study because it may contribute to the CT mechanism of the alternative pathway. Therefore, it can be supposed that this Trp extends the pathway and form a Tetrad. Moreover, this residue forms a π-stack complex with the other antenna chromophore of *PhrA*, a methenyltetrahydrofolate (MTHF) antenna, which harvests light energy to transfer it to FAD allowing the formation of FAD^*_{ox}. Furthermore, the mutant of this Trp induces a loss of this chromophore, hence it can be supposed that it has some relevant meaning for the functionality of the protein.[69]

In general, Site-directed mutagenesis experiments are used to assess the CT ability of a pathway. Thus, a hypothetical electron transmitter is replaced by a redox-inactive amino acid such as Phe or Ala (see chapter 6). This forms mutants of PCF members, which have a hindered CT pathway and show a totally blocked or a significantly reduced FAD photo-reduction.[34,68,70] The presence of an alternative pathway can be demonstrated by the persistence of the CT process when the classical Triad is interrupted.[77] The *PhrA* mutants contrast with the previously described behaviour. The absorption spectrum of FAD_{ox} presents a classical monoexponential decay in wild type (WT) *PhrA*, while it shows a transition between a slow initial decay and a very fast decay when a Trp from either pathways is mutated to Phe or Ala. These two components reduction patterns imply a slower secondary mechanism such as a conformational rearrangement. Therefore, it can be suggested that both pathways are required for the normal photo-reduction mechanism, which is characterised by a monoexponential decay of the oxidised FAD, FAD_{ox}, absorbance.[69]

In this chapter, the comparison and the interplay between the two Trp pathways is focused. It is considered that Trp336 participates in the alternative pathway, which is thus extended to a Trp Tetrad. The same unbiased non-adiabatic CT-QM/MM MD method is used to characterise the CT, which was already successfully applied to characterise CT in the last chapter 6, and to investigate the CT in DNA,[132,134,135,188] in PL,[16,17,26] in CRY[18] and also in organic materials.[130,136,189]

Computational results for the classical Trp Triad in PL and CRY were archived which match to the experimental measured reaction times. This approach scheme allows the direct simulation of the charge motion along the different redox cofactors without any bias, which unravels the relative stability of the different redox states and the possible backward CT mechanisms. It was shown that the electrostatic environment, provided by the protein structure and the solvent accessibility, generates a downhill picture for the CT along the Triad.[18,26] Such a stabilisation is required in the signalling functionality

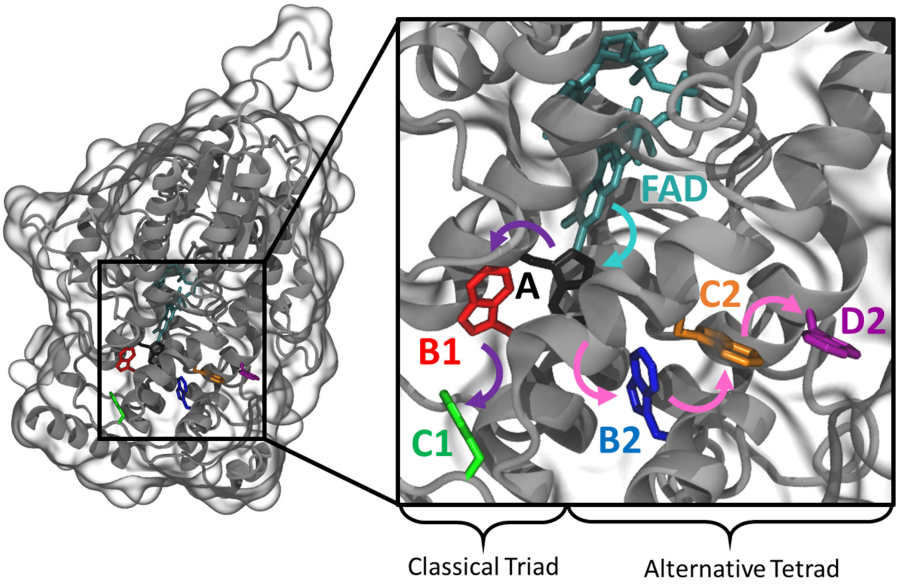

Figure 7.1.: PhrA structure (PDB:4U63)[69] and the charge migration within the two CT pathways: the classical path, Triad, with Trp384 (**A**, black), Trp361 (**B1**, red) and Trp308 (**C1**, green), the alternative path with Trp384(**A**), Trp318 (**B2**, blue), Trp367 (**C2**, orange), Trp336 (**D2**, purple) and FAD is coloured in cyan and shown explicitly. Coloured arrows represent the different CT: from FAD^{*}_{ox} to **A** (cyan), along the classical Triad (violet) and along the alternative Tetrad (pink). Hydrogen atoms are omitted for clarity and the picture is made with VMD.[145]

of CRY or to allow the protonation of FAD_{ox} which occurs at larger timescales (see for example Ref. 43 or chapter 9). Therefore, the influence of the solvent is also investigated because the plant CRY and class-III CPD-PL are related to a similar branch of the evolution of this protein family.

The main goal of this chapter is to establish the pertinence of the combination of the branching pathways at a molecular level. Thus, rates of the CT processes will be determined, which follow the FAD photo-reduction. Firstly, each pathway is considered separately to analyse their own characteristics and interactions with the protein environment. Afterwards, both branching CT pathways are included to the simulation protocol in order to define the interplay and correlation of different CT mechanisms in *PhrA*. The first molecular insight into the CT pathways and the experimental results by Lamparter et al. raised the following question which will be investigated in the next sections:

- How are the CT kinetics along the conserved and alternative pathway in *PhrA* and why could both pathway be required?
- Does the solvent influence the CT in the same fashion as in plant CRYI of *Arabidopsis thaliana*? If yes, are both pathways affected similar?
- Can the extended size of the quantum region be simulated accurately with the approach during the requested long time scale?
- Are structural reorientations induced by the formed separated charge states?

7.1. Computational details

7.1.1. MD Simulation

The crystal structure of *PhrA* (PDB-ID:4U63)[69] was used as starting structure. The construction of the simulation box and all the classical simulations were performed using the GROMACS 5.0.4 package[147,148] and the AMBER-SB99-ILDN force field.[149,150,190] The force field parameters for the different FAD redox states were previously described in Ref. 17,18 and chapter 6, 9 The atomic charges of the MTHF antenna chromophore are generated from restrained fitting on the electrostatic potential (RESP)[153,191] calculated at HF/6-31G*[154,165] using Gaussian09 software.[155] Other non-standard force field parameters were derived from the general AMBER force field (GAFF).[151,152] The pKa of titratable residues has been determined using PROPKA 3.0 program.[192,193] The protein was solvated in a $118\times118\times118$ Å^3 TIP3P water[159] cubic box. Three chloride anions were added to neutralise the system.

The equilibration of the solvated protein *PhrA* started with a minimisation step, followed by 100 ps MD in the canonical ensemble and 100 ps in the isothermal-isobaric ensemble (300K, 1 atm) using Nose–Hoover thermostat[112] and Parinello–Rahman barostat.[113] Then, 150 ns of production of MD simulations in the isothermal-isobaric ensemble (300K, 1 atm) were performed with a time step of 2 fs and covalent hydrogen bonds were fixed on a constant length by the use of the LINCS algorithm.[160]

7.1.2. QM/MM simulations

29 geometries from the 150 ns classical MD trajectory were extracted, which were separated by one nanosecond in 6 intervals: 25-29 ns, 35-39 ns, 71-74 ns, 96-100 ns, 121-125 ns and 146-150 ns. For the non-adiabatic CT-QM/MM MD simulations, the same procedure as in chapter 6.1 and Ref. 17,18,26 were considered.

The first photo-reduction, namely the reduction of FAD_{ox} to $FAD^{\bullet-}$ and the sub picosecond CT between the FAD^{*}_{ox} and the closest site **A** were excluded. Thus, it was assumed that they already happened and form the starting point for the further CT simulations. Therefore, FAD is described using classical force field on its negatively charge state, $FAD^{\bullet-}$ and the QM region contains the side chain of the Trp residues from the considered pathway(s). QM/MM boundaries, defined on the C_α-C_β bond, were treated using hydrogen link atoms. Taken together, three QM/MM partitions were used, depending on the considered CT cascade, therefore, the following QM regions were constructed:

- The first possible QM region concerns the classical Triad: containing only **A**, **B1** and **C1** side chains
- The second QM region corresponds to the alternative Tetrad: including **A**, **B2**, **C2**

and **D2**

- The third partition includes both pathways: site **A**, **B1**, **C1**, **B2**, **C2** and **D2**

All QM/MM simulations last 1 ns. 25 or 29 QM/MM charge propagation simulations along the individual pathways or the branching pathways were performed, respectively.

Additionally, QM/MM simulations in which the positive charge remains on one of the Trp residues were carried out for 1 ns, while the FAD cofactor is negatively charge. The difference between the site energy for each positively charged Trp and their neutral state corresponds to the stabilisation of the additional charge on each fragment, induced by the surrounding structures. This stabilisation is split into two components: the structures forming the CT-pathway, which are FAD and the six Trp residues, form one part of the studied interaction. The environment, which is the solvent and the protein excluding the six Trp residues and FAD, rises the second part of the interaction. The first contribution is obtained by calculating the site energy of the charged Trp along the previously obtained 1 ns simulation with all the atomic charges equal to zero except for the FAD cofactor and the Trp side chains. Then, the difference between the total energy, including all interactions, and the determined energy of just the CT pathway contribution, provides the contribution induced by the environment.

7.2. Charge transfer along the independent pathways

The Trp384, mentioned as site **A** is in a suitable position to give an electron to the FAD^*_{ox} and hence forms the first site of the studied CT pathways. Afterwards, the continuous CT can involve the following two branching pathways: the classical Triad, containing **A**, **B1** and **C1**, or the alternative Tetrad, which contains also **A** and further **B2**, **C2** and **D2**, explicitly presented Figure 7.1.

Firstly, the conformational space around the different Trp residues is sampled along a 150 ns classical MD. Thereafter, unbiased non-adiabatic CT-QM/MM MD simulations of the CT propagation along the classical Triad or the alternative Tetrad are performed. In a first set of simulations, the two pathways are considered individually in order to compute rates for the respective pathways. Starting from 25 different structures which are randomly extracted from the long classical MD simulation, direct CT simulations over the classical Triad are performed: only **A**, **B1** and **C1** compose the QM region in the simulations, therefore, no transfer to the alternative pathway or recombination with FAD are considered. Exactly the same procedure is applied to the alternative Tetrad, considering **A**, **B2**, **C2** and **D2** in the QM region.

The occupation of each Trp, which directly relates to its charge (0 means neutral site and 1 fully oxidised site), is evaluated as a function of time during the 1 ns QM/MM simulations. Then, this occupation is averaged on the 25 simulations to draw kinetic profiles, shown in Figure 7.2. The corresponding rate constants are listed in Table 7.1, determined from the time-dependence of the averaged occupations following a numerical fitting procedure described in Ref. 18 and Eq. 6.1 in chapter 6.

The kinetic profile for the classical Triad **A**-**B1**-**C1**, which is shown in Figure 7.2, describes a very fast CT cascade: **B1** is transiently occupied and the charge is stabilised on the well-solvated last site **C1** after 20 ps of simulation on average. A remarkably high rate constant of 1010.3 ns^{-1} for the forward CT from **A** to **B1** (7.1) is observed, which is about 5-times the **B1** to **C1** CT rate constant. This rate constant is about 10-times faster compared to the corresponding rate constant in CRYI of *Arabidopsis thaliana*, abbreviated in the further part as CRYI.[18] The slowest backward transfer from **C1** to **B1** reflects the stabilisation of the charge on the last Trp **C1** (7.1).

The same procedure is applied on the alternative Tetrad **A**-**B2**-**C2**-**D2**. The corresponding kinetic profile presents a charge stabilisation on **D2** after 60 ps on average, shown in Figure 7.2. The occupation increases on site **D2** and decreases on site **A** in a symmetrical way, whereas the occupation on site **B2** and **C2** stays very constantly low. The **C2** occupation slightly grows in the first 20 ps to get an averaged occupation of 10% which remains stable during at least 1 ns. The large standard deviation values in Figure 7.2 of the occupation of **C2** and **D2** denotes different final distribution of the charge along this Tetrad.

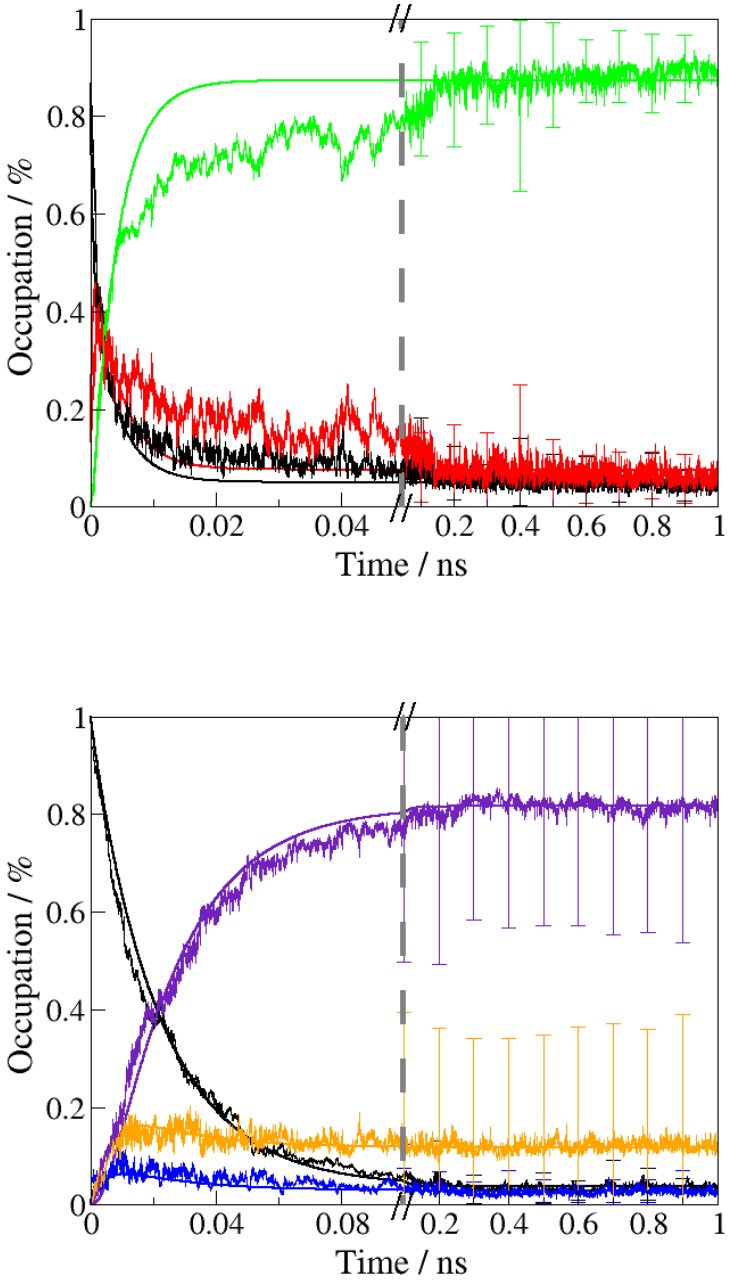

Figure 7.2.: Time dependent evolution of the averaged charge occupation along the classical Triad and the alternative Tetrad in *PhrA*: the upper picture shows the averaged population of the charge in the QM/MM simulations, having just the Triad included (black line: **A**; red line: **B1**; green line: **C1**) and the lower picture shows the averaged population of the charge in the QM/MM simulations, having just the Tetrad included (black line: **A**; blue line: **B2**; orange line: **C2**; purple line: **D2**). A timescale change is represented by the grey dashed lines to highlight a zoomed in picture of the fast CT processes in around 0.05 ns for the Triad and 0.1 ns for the Tetrad. Error bars denote the standard deviation of the occupations each 0.1 ns. The charge is in both sets of simulations placed on **A** at the starting point. The fitting curves from the kinetic model are also shown using solid lines in the same relative colours.

Table 7.1.: Rate constants in ns^{-1} for the different CT transfers between Trp in *PhrA*. The first subscripted letter indicates the positive charge donor and the second the acceptor. Each line corresponds to a different QM part: only **A**, **B1** and **C1** for the "classical Triad"; **A**, **B2**, **C2** and **D2** for the "alternative Tetrad" and finally all Trp residues for "both" pathways.

	k_{AB1}	k_{B1A}	k_{B1C1}	k_{C1B1}	k_{AB2}	k_{B2A}	k_{B2C2}	k_{C2B2}	k_{C2D2}	k_{D2C2}
Classical Triad	1010.3	778.3	219.6	20.0	-	-	-	-	-	-
Alternative Tetrad	-	-	-	-	48.6	56.6	1195.5	318.0	176.7	25.4
Both	738.6	601.6	435.1	71.4	26.9	46.0	398.4	196.5	61.1	7.7

In nearly all the QM/MM simulations, the positive charge is stabilised on **D2**, supporting the hypothesis of a participation of this Trp to the alternative pathway. In just two simulations, the positive charge is stabilised on **C2**, while a conformational change is observed: the distance between the centres of mass of **C2** and **D2** increases from 7.6 to 10.0 Å. Hence, the electronic coupling is diminished by two orders of magnitude, thereby blocking the transfer to site **D2**. The CT between **C2** and **D2** is thus coupled to conformational changes, which has already been mentioned for the transfer between Trp **B1** and **C1** in the Triad of CRYI.[18]

The rate constant of the first transfer between **A** and **B2** is around 50 ns^{-1}, and therefore 20-fold slower than between **A** and **B1**, presented in Table 7.1. Besides, the corresponding backward transfer from **B2** to **A** occurs on the same timescale. On the contrary, the second CT between **B2** and **C2** happens on a similar timescale as the **A-B1** (thousand ns^{-1}). This combination of a slow oxidation of **B2** and a fast oxidation of **C2** leads to an unoccupied **B2** site. The CT from **C2** to **D2** is slower, with a rate around two hundreds of ns^{-1}, but still one order of magnitude higher than the corresponding backward transfer. This difference between $k_{C2\text{-}D2}$ and $k_{D2\text{-}C2}$ attests to the stability of the charge on the last Trp **D2**.
In the alternative Tetrad, the first CT constitutes the kinetic limiting step while the charge is then quickly transferred to **D2**, in agreement with the symmetric time-dependent profiles of **A** and **D2** occupations, shown in Figure 7.2.

The individual analysis of the QM/MM simulations provides further details about the relative stability of the charge on the last residue of each pathways. The frequency of the backward transfers from the last residue **C1** or **D2** to **A** is determined. On average, 8 backward transfers occur during the classical Triad trajectories and are randomly distributed in the 1 ns simulation time.
On the contrary, in the alternative Tetrad, the number of backward transfers is 2.4 on average, and they happen only at the beginning of the simulations: no backward transfer are observed after 250 ps. Most of the simulation time, the positive charge stays localised on **D2** or **C2**, which decreases the probability of a charge recombination on

FAD. This observation suggests that the conformation of the close environment rapidly adjusts to stabilise the positive charge on **D2**.

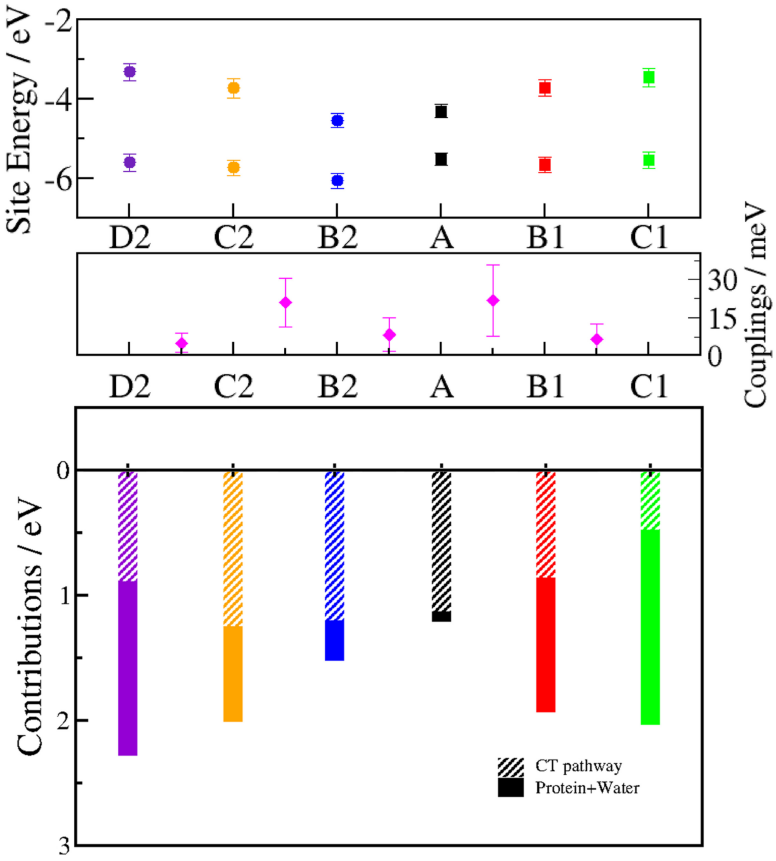

Figure 7.3.: Site energies of the different Trp residues of the Triad and Tetrad: the colouring scheme is as usual: Triad is coloured in black (**A**), red (**B1**) and green (**C1**) while the Tetrad is coloured in blue (**B2**), orange (**C2**) and purple (**D2**). The energy of the neutral state is shown in the top graph, lower values and the oxidised state is indicated by the higher energy values (top graph, higher values). Energy values are averaged along 1 ns QM/MM MD trajectory and the error bars denote the standard deviation of the site energies. The energy gap between the two states (neutral-oxidised) is divided into two contributions for each Trp: the interaction with the other Trp residues and with $FAD^{\bullet-}$ (diagonal lines) and the remaining system which is the protein and solvent (solid bar).

Though isoenergetic in gas phase, the different Trp residues in *PhrA* have different ionisation potentials (IP), based on their heterogeneous environment of the solvated protein. Figure 7.3 presents the site energy of each Trp residue with respect to the neutral and the oxidation states $IP \approx -HOMO \approx -site\ energy$. The energy of the neutral and oxidised Trp residues are averaged over 1 ns simulation. The energy of **B2** is slightly higher than **A**, therefore, the observation of the slow rate constant of **A**-**B2** transfer is consistent with the energetics. Conversely, the large energy differences between **A** and **B1** (about 0.6 eV) or **B2** and **C2** (about 1 eV) correlates with the very fast forward CT along the Triad. For both pathways a downhill energy profile is observed, which was previously highlighted in CRYI, class-I CPD-PL of *E. coli*[17,18] and in *PhrB* shown in

chapter 6.[26]

To study the stabilisation of the positive charge at the different Trp residues in more details, the energy difference is decomposed into two contributions, which are also presented in Figure 7.3: (I) the interaction with CT-pathway, containing just the six Trp side chains and $FAD^{\bullet-}$, and (II) the interaction with the remaining protein and the solvent molecules. A localised positive charge on **A** is stabilised by the interaction with the negatively charged $FAD^{\bullet-}$ strongly. As expected, the contribution of the FAD-Trp interaction decreases when the FAD-Trp distance increases. Therefore, the stabilisation of the oxidised state of **B1**, **C1** and **D2** is mainly based on interactions with the protein and the solvent molecules. However, the contribution from the CT-pathway and FAD constitute the major part of the stabilisation of **B2** and also **C2**.

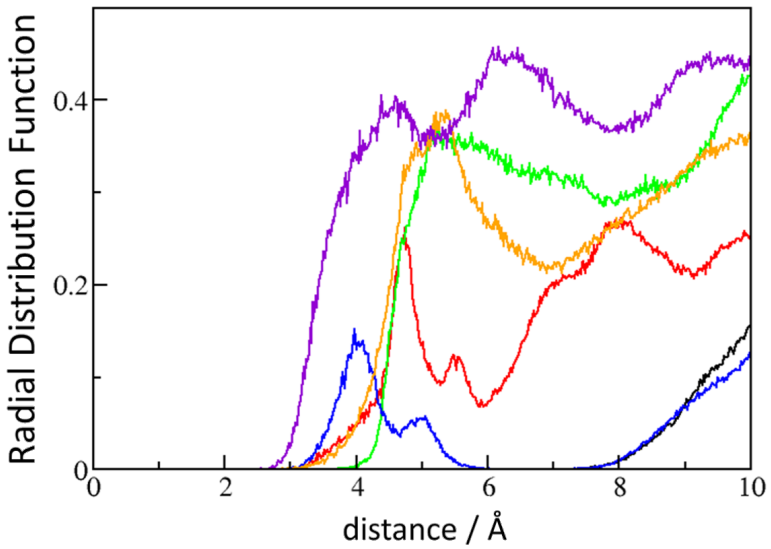

Figure 7.4.: Radial distribution function of water molecules around the center of mass of the different tryptophans in the branching pathway: Black line: **A**; red line: **B1**; green line: **C1**; blue line: **B2**; orange line: **C2**; purple line: **D2**.

The radial distribution function of water molecules around the sites is shown in Figure 7.4. The sites of the classical Triad presents an increasing solvation from **A** to **C1**, comparable to the water molecules distribution around the Trp Triad in CRYI.[18] Much less water molecules interact with **B2** than with **B1**, furthermore, a look at the structure reveals that **B2** is surrounded in a mostly hydrophobic environment, which explains its relatively high site energy. **C1** and **C2** have a similar solvation, and **D2** is in direct contact with the bulk water.
For other PLs and CRYI it was shown that the typical downhill profile for the CT results from better solvation of the residues. Therefore, sites which are more exposed to the bulk solvent showed a stronger stabilisation of the positive charge, favouring the stabilisation.[17,18,26]

Further, the radical is stabilised by the deprotonation of the last Trp residue of a pathway,

leading to the neutral radical form, which is energetically favoured. The deprotonation of the homologues of **C1** occurs in 350 ps in a class-II CPD-PL,[194] 200 ns in CRYI,[195] 300 ns in class-I CPD-Pl of *E. coli*[196] and 2.5 μs in *Xenopus laevis* PL.[183] The sub-nanosecond deprotonation in the class-II CPD-PL is explained by the presence of an acidic wire in the site **C** neighbourhood, which contains a water cluster, an aspartate and a histidine. In *PhrA*, several water molecules can also constitute a hydrogen bond wire around **C1**, however, no acidic residue is available, which could support the deprotonation of the positively charged **C1** in the same fashion.

The **C1** pocket is relatively similar to the Trp324 pocket in CRYI, except for the presence of a salt bridge close to **C1**. Considering these structural comparison, it can be expect that the deprotonation of **C1** is faster than in CRYI, but not as fast as in the class-II CPD-PL. In absence of any experimental value, the competition between the transfer to the alternative Tetrad and the deprotonation of **C1** remains conceivable, but it is not considered in the simulations. In most of the simulations, the nitrogen atom of **D2** is in hydrogen bond interaction with the backbone of Thr319. However, some **D2** side chain rotations are observed leading to the complete solvation of the indol group, and therefore the Trp deprotonation is supported and can happen faster.

In summary, the results on the independent CT pathways confirm the relative importance of the different Trp residues from both pathways. Once **A** is oxidised, the charge moves faster on the classical Triad than on the alternative Tetrad and the positive charge is stabilised on **C1**, following a downhill mechanism. A similar profile is observed for the alternative Tetrad, but the CT between **A** and **B2** is two orders of magnitude slower than the **A-B1** CT.
However, the stabilisation of the charge is stronger on the last Trp **D2** of the alternative Tetrad than on the site **C1** of the classical Triad. If the distance between **C2** and **D2** increases and strongly disfavours a CT, the good solvation of **C2** can allow a significant stabilisation of the charge on the alternative pathway. Additionally, the backward transfer is energetically impeded by the high cost of a the reoxidation of **B2**. Consequently, the simulations draw a distinction between a "kinetically" favoured Triad **A-B1-C1** and a "thermodynamically" favoured Tetrad **A-B2-C2-D2**, having their own advantages and drawbacks.

7.2.1. Charge transfer along the branching pathways

Site-directed mutagenesis of Trp residues from both pathways lead to a complex photoreduction process which contrasts with the monoexponential decay observed after excitation of FAD_{ox} in the WT and suggests an important role of the combination of two CT pathways.[69] The simulations, reported above, neglect the interplay between both pathways by considering just one. Thus, non-adiabatic CT-QM/MM MD simulations are performed to study the correlation between the CT along both pathways, including **A**,

B1, **C1**, **B2**, **C2**, **D2** in the QM region. 29 individual 1 ns QM/MM simulations are performed, which have randomly selected starting structures and an oxidised site $\mathbf{A}^{\bullet+}$ and a classical $FAD^{\bullet-}$. Similar to the previous parts, averaged occupations of each Trp over the different simulations are used to draw the kinetic profile, which is presented in Figure 7.5.

During the first 10 ps, the charge promptly goes from **A** to **C1**, thus, maximum of **C1** occupation (about 70%) is observed after 20 ps. Subsequently, the **C1** occupation slowly decreases whereas the **D2** occupation conversely increases. After 1 ns, the final occupation of **C1** and **D2** is between 40% and 50%, respectively. Because of the larger stabilisation of the charge on the alternative Tetrad, a complete transfer of the charge on **D2** can be expected after several nanoseconds. The occupations of the intermediate Trp residues **B1**, **B2** and **C2** remain very low as well as the **A** occupation. Actually, in 27 of the 29 simulations, the charge moves to **B1** first and in two remaining trajectories, the charge is directly propagated along the alternative Tetrad. Transfers causing crossings between pathways occur at low frequency (at maximum 3 crossings per simulation), but they lead to a 40-60% distribution of the charge on the two pathways: after 1 ns, 12 trajectories present an oxidised $\mathbf{C1}^{\bullet+}$ and 17 an oxidised $\mathbf{D2}^{\bullet+}$.

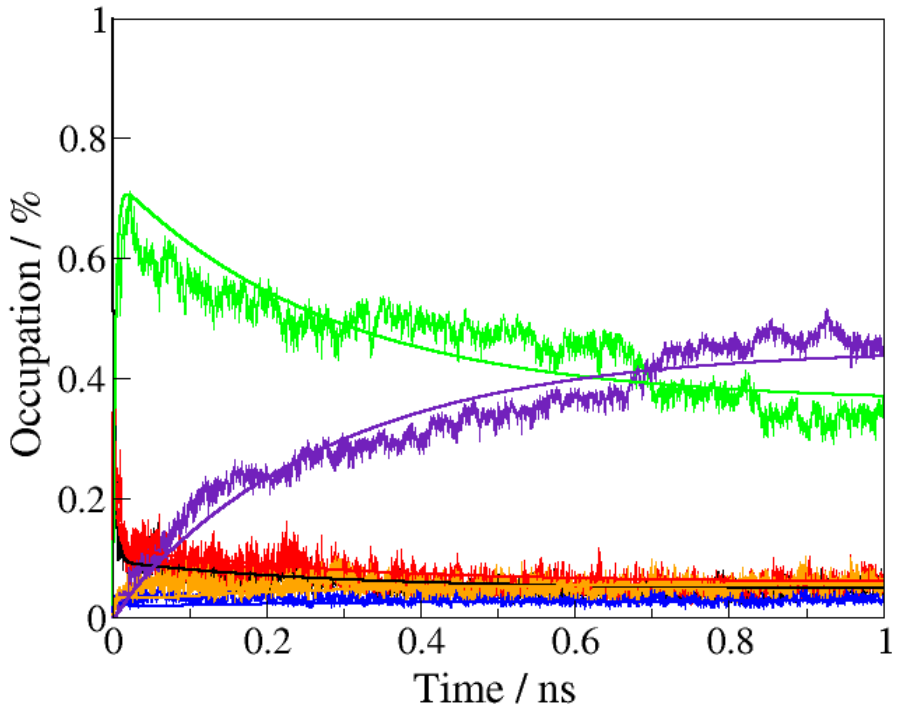

Figure 7.5.: Time dependent evolution of the averaged charge occupation along both pathways in *PhrA*: Black line: **A**; red line: **B1**; green line: **C1**; blue line: **B2**; orange line: **C2**; purple line: **D2**. The charge is on **A** at the starting point of the simulation. The fitting curves from the kinetic model are also shown by a solid line, using the same relative colors.

This equal distribution of the ending Trp residues of the Triad and Tetrad may also look like an artificial failure of the used propagation method of the SE. The applied

Ehrenfest propagation is used which enforces a delocalisation of the charge carrier over the whole system. This was already proofed by Kubaš for the CT in DNA and explained in details in Ref. 188 However, the studies of the other PL and CRY showed no artificial delocalisation, which is probably compensated by the individual environment of the protein, forming different interaction patterns with the sites.

Therefore, the question arises about a delocalisation of the charge along the more similar Trp residues **A**, **B1** and **B2**, which could induce a spreading of the charge over the whole system. The occupation of all sites during all QM/MM simulations with the two pathways are analysed and shown in Figure 7.6. This shows that the charge is not spread over the whole system, occupying the endings of the pathways. However, just site **C1** or **D2** is occupied for around 80-90%. Thus, the charge is not transferred *via* an artificial delocalisation between the Triad and the Tetrad, additionally, most of the time the charge stays localised at one site and travels in this constitution along the pathways.

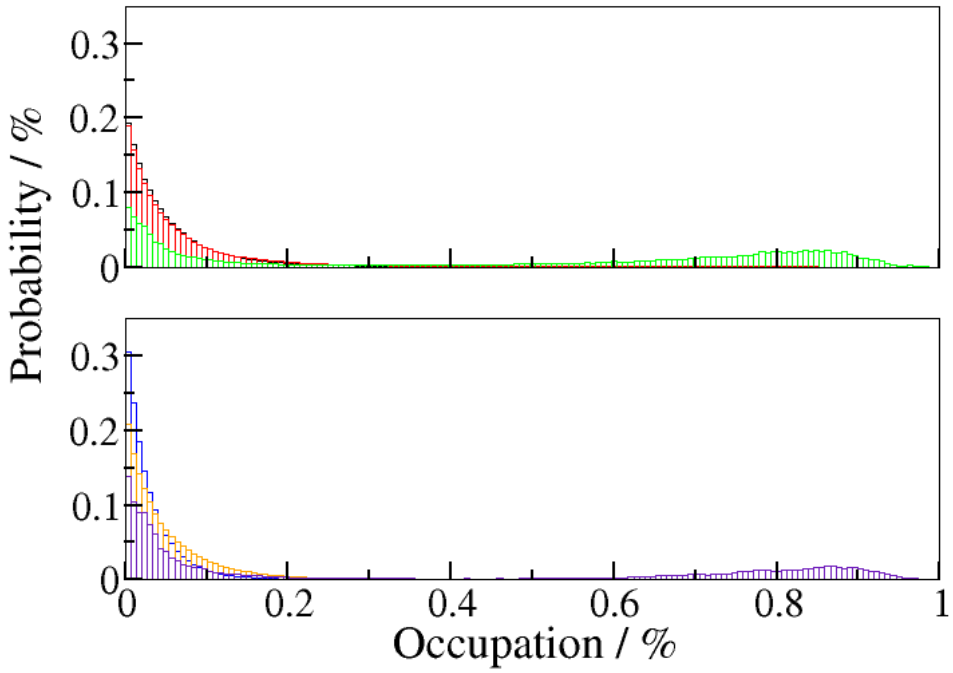

Figure 7.6.: Averaged population of the individual sites: the same colouring scheme as in Figure 7.5 is applied and the weighted histogram of the charge population of the individual sites are presented. The 29 QM/MM trajectories are used for the average.

The rate constants, presented in Table 7.1, are influenced by the extension of the QM region to the six Trp residues. As expected, a decrease of the rate constants occurs, involving site **A** because the Y-shaped pathway induces a competition between **A**-**B1** and **A**-**B2** transfer. Obviously, the classical Triad has the better kinetic parameters, which is enabled by the stronger electronic couplings, allowing a fast transfer of the

positive charge. However, the rate constant of the second CT, from **B1** to **C1**, is lower than the backward CT rate from **B1** to **A**. Thus, the charge fluctuates along the classical Triad more rapidly than along the alternative Tetrad. On the contrary, if the first transfer from **A** to **B2** is slow, the second transfer from **B2** to **C2** is ten-fold faster than the backward transfer from **B2** to **A**. Therefore, the barrier for backward CT in the alternative Tetrad is higher than in the classical Triad. Based on that, the charge is more stabilised on **C2** or **D2** than on **C1**, shown in Figure 7.3.

7.2.2. Environment around FAD

Furthermore, QM/MM simulations are analysed in which the positive charge is constrained on one site and $FAD^{\bullet-}$ is present. The starting structure is taken from the 150 ns classical MD simulation where the FAD_{ox} is simulated and the Trp remained neural. Therefore, manually the CT are performed and constrained at each Trp to see their interplay with the environment. The conformational changes around each redox partner are investigated, induced by the charge separation and the positive charge propagation.

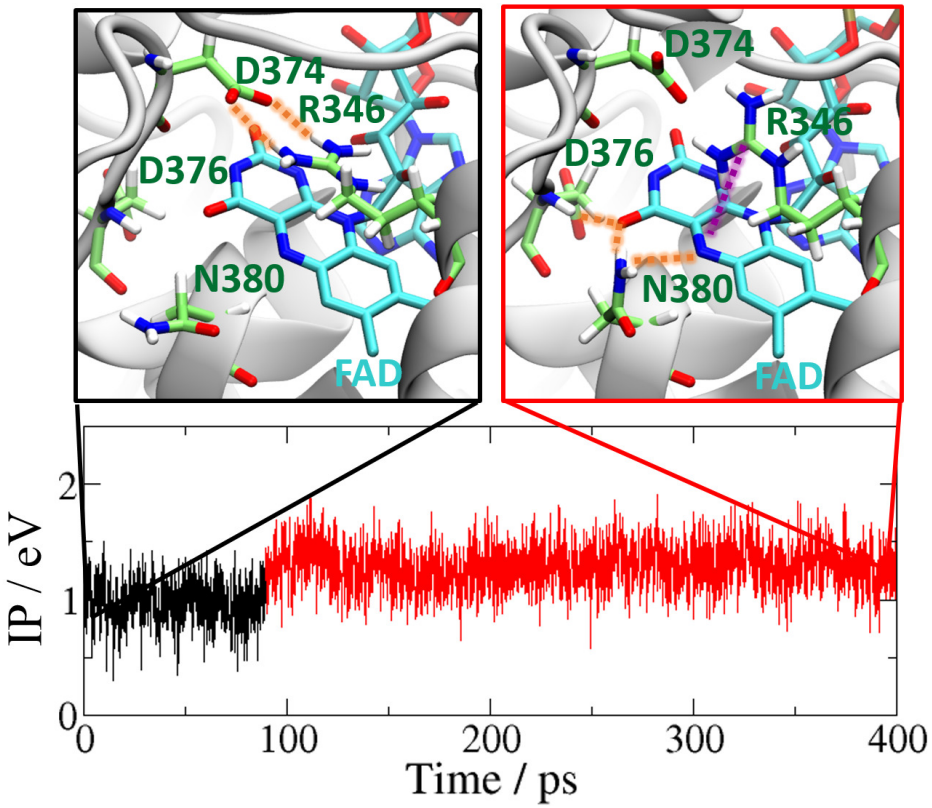

Figure 7.7.: Fluctuation of the ionisation potential of $FAD^{\bullet-}$ when the charge is fixed on **C1**: the starting and the last conformations (upper pictures) of the FAD pocket illustrate different modifications of the environment occurring after the FAD photo-reduction. Orange dashed lines indicate the hydrogen bond and purple dashed line the π-cation interaction. Hydrogen atoms of FAD have been deleted for clarity.

After almost 100 ps, the $FAD^{\bullet-}$ IP increases by 0.2-0.4 eV, making the reoxidation of FAD harder, which is shown in Figure 7.7, i.e., the energy of the hole on FAD is increased by environment reorganisation. Several examples of this reorganisation are represented in Figure 7.7: the rotation of Asn380, the rotation of the guanidinium group of Arg346 from π-stacked conformation to π-cation interaction with FAD, the formation of a hydrogen bond between Asp376 backbone and FAD.
Thermodynamically, this will not have a big impact on the driving force of the charge recombination, however, the reorganisation, in a Marcus picture, would change the curvature of the diabatic states and may thereby introduce a small kinetic barrier. The electronic coupling between **A** and FAD_{ox} or $FAD^{\bullet-}$ does not change, thus, it is not affected by these structural reorientations. Overall, a recombination due to backward transfer along the classical Triad seems likely. Then, a competition between the charge recombination on FAD and the transfer to alternative pathway arises. Once the charge is on **C2** or **D2**, nevertheless, backward transfer is strongly inhibited by the large energy gap in the **C2**-**B2** transfer.

The mechanism observed in the QM/MM simulations can be described by three steps, resumed in Figure 7.8: (I.) the charge is rapidly transferred to **C1** through **B1**; (II.) the environment stabilises the negative charge on the isoalloxazine ring of $FAD^{\bullet-}$, and the charge is transferred back to **A**; (III.) the charge is stabilised on **D2**. The fast transfer to **C1** avoids the immediate charge recombination between **A** and $FAD^{\bullet-}$. However, backward CT are frequent, rising the risk of FAD reoxidation due to charge recombination when **A** is again transiently positively charged, $\mathbf{A}^{\bullet+}$.
Based on that, two mechanisms counteract the charge recombination: the deprotonation of positively charged **C1**, which happens likely in several nanoseconds or the transfer of the positive charge to **B2**, which can happen in about several hundreds of ps, shown in the non-adiabatic CT-QM/MM MD simulations. A reorganisation of the environment around the FAD chromophore seems to follow the redox state modification of the isoalloxazine ring. On the contrary, the hydrophobic pocket of **B2** appears mostly insensitive to the redox state of **B2**, highlighted by the sources of the stabilising interactions, which is presented in Figure 7.3. Therefore, the charge can be transferred from **A** to **C2** through **B2** without any outer sphere reorganisation. Then, the stabilisation of the positive charge on **C2** or **D2** prevents any charge recombination finally.

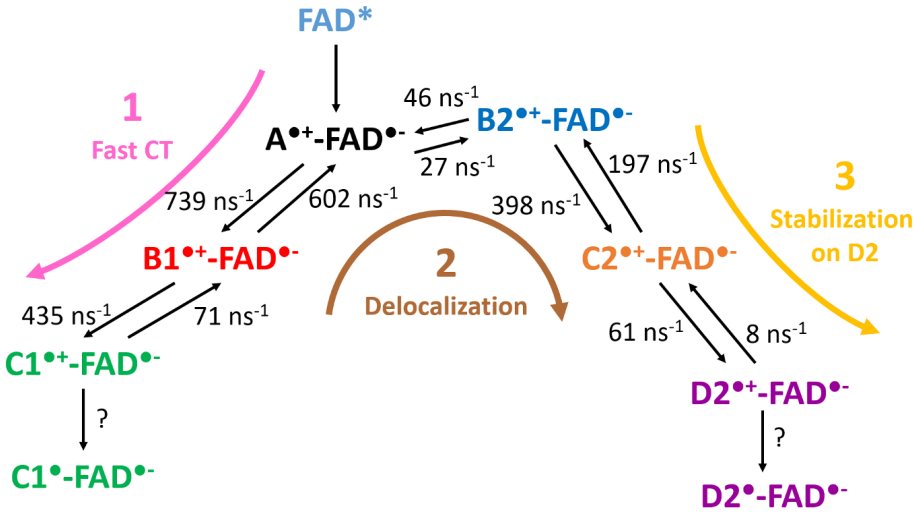

Figure 7.8.: Scheme summarising a possible mechanism involving both kinetic and thermodynamic pathways in the CT transfer in *PhrA*: the numbers correspond to the reaction rates, presented in Table 7.1, the individual sites are coloured with the introduced scheme, the pink arrow illustrates the kinetically favoured path and the yellow one the thermodynamically one.

7.3. Conclusion & Discussion

PL and CRY are able to efficiently photo-reduce FAD^{*}_{ox} into $FAD^{\bullet-}$. Several pathways of aromatic residues participate in the CT from the FAD to the solvent. *PhrA* possesses two related CT pathways, starting from the same Trp384. One of these pathways is called the classical Triad with reference to the homologous pathway, firstly described in *E. coli* PL. The second one represents an alternative pathway which was also previously described in other proteins such as CRYI of *Arabidopsis thaliana*.[71,77,197]
Such a pathway can contribute to the photo-reduction of FAD, if the CT along the classical Triad is hindered. In *PhrA*, however, the FAD photo-reduction *in vitro* is altered by mutations on both pathways.[69]

This computational study provides a new insight on the respective role of this alternative pathway and the interplay with the classical conserved Triad. First of all, a fourth tryptophan is detected and the calculations strongly suggest that it is involved in the CT along the alternative pathway. It shows the lowest IP, and therefore it is the most convenient candidate for a good stabilisation of the positive charge.
However, the classical pathway presents the highest rate constant for the first CT from the first Trp **A** to the second of the CT cascade. The comparison between the rate constants of the CT along both pathways and the IP of the different Trp residues forms the following descriptions: the classical Triad can be seen as the "kinetic" pathway and the alternative Tetrad as the "thermodynamic" pathway. The unbiased non-adiabatic CT-QM/MM MD simulations reflect these different behaviours and in most of the 1 ns trajectories the charge goes to the classical Triad in few picoseconds. From time to

time, a transfer to the alternative Tetrad occurs and then the charge is stabilised on the last Trp **D2**. After 1 ns, a similar probability to have the charge on **C1** or **D2** occurs, but the kinetic profiles and the energetic values suggest a final stabilisation of the charge on **D2** after several nanoseconds. In about 100 ps, a reorganisation of the FAD pocket leads to a stabilisation of $FAD^{\bullet-}$ due to structural reorientations of surrounding amino acids.

The ubiquity of an efficient CT pathway in all members of the PCF underlines the importance of the photo-reduction process leading to a stable radical state, $FAD^{\bullet-}$+ $Trp^{\bullet+}$. Several strategies have been followed by the evolution to maintain the positive charge from the FAD: the fast CT to the first and the second Trp avoids the direct and energetically favoured charge separation-recombination mechanism, which occurs in a solvated U-shape FAD;[72,198] the common downhill process due to the environment creates a stable charge separation state, $FAD^{\bullet-}$- $Trp^{\bullet+}$;[18,26] the fast deprotonation of the last Trp stabilises its oxidation;[194] or in the case of *PhrA*, the presence of an alternative pathways. The combination of both CT pathways in *PhrA* ensures a fast and stable FAD photo-reduction. This strategy can be also seen by many CRYs and PLs containing homologues of the branching pathway of *PhrA*.

8. Benchmark of the multi-scale charge carrier propagation method

Applications of the multi-scale method, non-adiabatic CT-QM/MM approach, in the biological systems showed that this approach is able to characterise the CT processes in additional member of the studied protein family (Photolyases/Cryptochromes), presented in the previous chapters 6, 7.
In addition to the application in the biological systems, this method was extended and used to reproduce the experimentally measured charge carrier mobility of organic semiconductors (OSC).[130,136] An Ehrenfest propagation method was successfully used to gain qualitative insight into the mechanisms of the CT processes in biological systems and OSC. The treatment of the charge carrier evolution of the system on a single potential energy surface (PES) using this propagation approach also showed an artificial overdelocalisation of the charge carrier along the system. No included force addresses a relocalisation of the charge. Therefore, the charge will spread along the system, when the nature of the system does not offer barriers between the states. This problem was observed in the simulation of CT process in DNA using the introduced multi-scale method. Simulations of the CT along a homogeneous poly-adenine DNA strand showed this artificial spread of the charge, while simulations of an adenine and guanine DNA strand did not reveal this issue. This is based on the different energy levels of the DNA bases[133,188] and also holds for proteins which offer a different environment to in- or decrease the energy levels of the molecules, which are involved in the CT process (explicitly discussed in chapter 6,7 and Ref. 16–18).
An alternative most widely used method for non-adiabatic molecular dynamics is Tully's fewest switches surface hopping (SFH) propagation[140] and many other further derived SFH schemes. In contrast to the propagation on a single PES in the Ehrenfest method, the SFH methods allows classical trajectories to hop stochastically between PESs, governed by the comparison of a computed transition probability with a random number (for the details of see Section 5.5).
The reported results showed that the charge spreads in a homogeneous environment even if its polar (DNA) or apolar (crystalline OSC). This problem is addressed in this chapter, and therefore a SFH propagation is implemented in the local version of the Gromacs package, used for the multi-scale method. A SFH implementation was already used in a branched version of this multi-scale approach (with an addition extension to TD-DFTB) and obtained good results for an exciton transfer. The same SFH approach

will be included into the non-adiabatic CT-QM/MM MD method, to study the performance for CT in an OSC derived test system. The methodology of the implemented SFH and the taken assumptions are explained in section 5.5 and in Ref. 143,144. This shall increase the accuracy of the method, which can allow to increase the transferability of the multi-scale method. Thus, further unknown systems can be examined to support the experimentally difficult study of CT in e.g. proteins, OSC. Furthermore, the established Ehrenfest propagation will be benchmarked using a test system, therefore, the same assumptions are considered which are explained in more details in Ref. 130,132,133,188 and briefly summarised in section 5.5.
In a recent study, Blumberger and co-workers developed a similar multi-scale method, which is also an non-adiabatic fragment orbital (FO) MD simulation scheme using a SF propagation, to simulate charge carrier transport in the poly-ethylene OSC.[126,199,200] Their developed method predicts a transition between a band-like charge transport mechanism at high electronic couplings to a thermally activated hopping mechanism at low electronic couplings. They created a test system derived from a chain of ethylene molecules and changed their properties to enable a benchmark of the CT process, with respect to the applied properties. Based on the similarity between both multi-scale methods a similar test system is used in this work to benchmark both propagation methods.

8.1. Construction of the test system

The OSCs based on the anthracene and its derivatives have been extensively studied by experiments.[201,202] Therefore, it was also used to prove the multi-scale method, which was extended by Dr. Heck and Dr. Kranz to investigate the OSC.

The geometry of the crystalline anthracene orientates the single molecules in three different motives which correspond to the different axes of the crystal (a, b and c). This stacking of the molecules creates symmetric orientations which rise completely different electronic couplings between the molecules. Different mobilities and a corresponding travelling speed of the charge carrier belongs to the individual orientations. The extended multi-scale method was able to reproduce the charge carrier mobility of selected crystalline systems, such as anthracene. Additionally, the observed temperature dependent decrease in the charge carrier mobility of organic crystalline systems was reproduced correctly.[130,136]

Based on the achieved promising results for the OSC, an anthracene crystal is used to create the test system, which is comparable to the system used by Blumberger and co-workers to benchmark their method, briefly described in the following. They used a constrained chain of ethylene molecules, which are called ethylen-like because just the vibration along the C–C bond is possible. The CT is determined by two important

properties: the electronic coupling between two molecules and the relaxation energy, λ, of the molecule. The parameters of the force field are modified to establish different values for λ, with respect to the charge state of the ethylen-like molecules.
Compared to the artificial ethylene chain a complete anthracene crystal is used for the simulation in this study. Therefore, no unphysical constrains, are required and the QM region will be placed in the middle of the crystal, orientated along b-direction and the surrounding molecules will constrain the chain for the charge propagation naturally. In Figure 8.1, the used pathway is visualised. The chain of grey anthracene molecules, being around 4 Å apart, forms the pathway for the charge. The position of the charge is indicated with red spheres in the centre and the charge is spread over five molecules, as an example of the slight delocalisation of the charge carrier in these systems, which is also defined as polaron.[15] The simulation is focused on a transfer of a positive charge along this pathway.

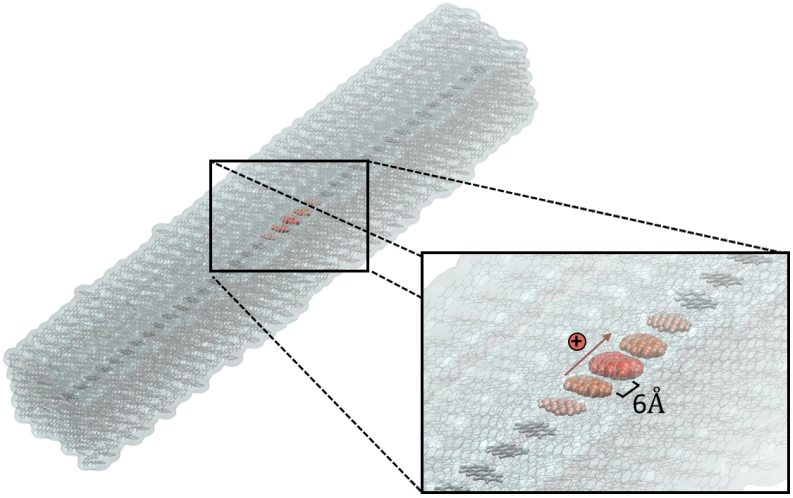

Figure 8.1.: Test system for the propagation benchmark: the grey line of anthracene molecules determines the QM region while the remaining molecules are presented with smaller sticks. The whole system is presented and a zoomed in view introduces the stacked chain of molecules. The position of the charge is visualised with the red spheres in the middle of the zoomed in view (right part). The travelling direction of the charge is indicated by the arrow.

In the anthracene-based test system, the CT will be studied while the ratio between λ and the coupling is changed. A correct propagation allows a transition between a band-like transfer (high coupling) to a thermally activated hopping mechanism (low coupling) for the CT.
The reorganisation energy is directly calculated from the non-adiabatic CT-QM/MM MD simulation derived by the FO-DFTB calculation. The difference between the energy of the neutral and charged molecule can be used to construct λ. The obtained value for λ agrees with calculations using DFT/PBE (explicitly and discussed in Ref. 130). The calculation of the coupling was benchmarked in earlier studies extensively, and a constant scaling parameter of 1.54 was achieved because the FO-DFTB calculation

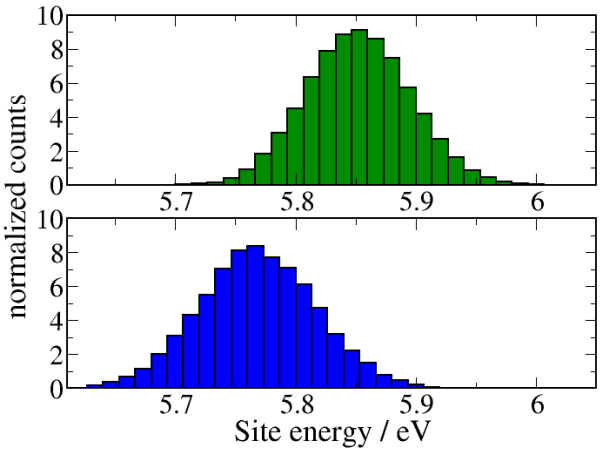

Figure 8.2.: Site energy of anthracene using the FO-DFTB approach: the upper graph shows the distribution of the site energy for a neutral anthracene molecule (peak of the site energy is around 5.85±0.43 eV, neutral). The lower graph shows the distribution of the site energy for a anthracene carrying a localised positive charge (peak of the site energy is around 5.77±0.48 eV, charged).

constantly underestimates the coupling.[168] In the same fashion, the coupling will be scaled to reduce the ratio with λ.

Thus, a different transport regime for the transfer will be established. Table 8.1 lists the different used values for the electronic coupling, with respect to the scaling parameter.

In Figure 8.2, the site energy and of the neutral and charged anthracene molecules are visualised. The site energy difference between the charged and neutral state is 83.6±4.5 meV (site energy$_{\text{neutral}}$=5.85±0.43 meV, site energy$_{\text{charged}}$=5.77±0.48 meV).

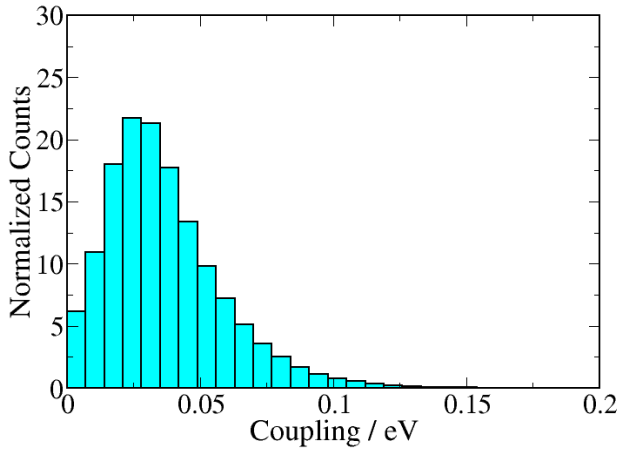

Figure 8.3.: Distribution of the electronic coupling between two of anthracene molecules using the FO-DFTB approach. The peak of the distribution is at 36.68±22.19 meV

The distribution of the electronic coupling is presented in Figure 8.3 and the peak of

the distribution is located at 36.68±22.19 meV which is less than half of the reorganisation energy and underestimates the coupling calculated with DFT (47.9 meV) even with the upscaled value. The scaling parameter of 1.54 is determined using a large set of molecules but could be individually optimised to study the properties of a system of interest.
The interesting transition between the transport regimes is resolved in the switching temperature dependency of the charge transfer mobility, which is linearly decreasing for band-like and parabolic increasing for hopping with respect to an increasing temperature. Therefore, the experimental measured mobilities will be used as reference and the following question which will be explained in this chapter:

- How does the Ehrenfest propagation treat the simulation in the low coupling regime?
- Does the implemented SFH propagation improve the multi-scale method?
- Can both propagators reproduce the transition between the band-like and hopping regime?

8.1.1. Methodology about charge transfer in OSC

A brief introduction into the CT in OSC is presented in the following. The charge carrier mobility can be derived from the diffusion motion of the charge according to the Einstein-Smoluchowski equation

$$\mu = \frac{eD}{k_B T} \tag{8.1}$$

which contains the elementary charge, e, the Boltzmann constant, k_B, and the diffusion constant D as

$$D = \frac{\langle \Delta x^2(t) \rangle}{nt} \ . \tag{8.2}$$

The diffusion constant relies on the dimensionality, n, of the system (2, 4 or 6 for 1-, 2-, 3-dimensional) and the time t. The mean-squared displacement (MSD) of the charge, $\langle \Delta x^2(t) \rangle$, is defined as

$$\langle \Delta x^2(t) \rangle = \sum_A (x_A(t) - x_0(t))^2 p_A(t) \ . \tag{8.3}$$

Where x_a is the centre of mass of molecule A and p_A is the occupation of the charge in molecule A. x_0 is the position of the charge at t=0 . This can be directly distinguished during the non-adiabatic CT-QM/MM MD simulation as $p_A(t) = |a|^2$, where a is the coefficient of the wave function in the used FO basis (see section 5.5 for more details).

Furthermore, the mobility is temperature dependent and described by the following

power-law

$$\mu(T) = C \cdot T^{-n} \tag{8.4}$$

The transfer along the different crystal axes showed rising values for n in the previous anthracene simulations, thus, the multi-scale method can treat transport regimes sensitively with respect to the electronic coupling.

8.1.2. Simulation details

The test system was constructed out of an anthracene crystal, containing 2000 anthracene molecules. A 10x40x5 crystal unit cell was used taken from Ref. 203. The GAFF force field parameters[151,152] were used for the bonded and Lennard-Jones interactions for the anthracene molecules. The atomic charges were calculated by restrained fitting on the electrostatic potential (RESP)[153,154] at HF/6-31G*[154,165] level with Gaussian09 package.[155] Periodic boundary conditions with particle-mesh Ewald electrostatics were used in all simulations. Anthracene was simulated at 300, 250, 200, 150 and 100 K and sets of the simulations for different electronic couplings for each temperature were performed.
Equilibration of the system started with a minimisation step, followed by 100 ps MD in the NVT ensemble and 100 ps in the NPT ensemble. 1 ns of production NPT MD simulations were performed afterwards using the Nosé-Hoover thermostat[112] to keep a constant temperature at 300 K and Parinello-Rahman barostat[113] to keep the pressure at 1 atm. The non-adiabatic CT-QM/MM MD scheme was implemented in a local version of the GROMACS 4.6. package (same version than in chapter 6 and 7). The QM region contains a chain of 30 molecules and the charge was located in the middle molecule of the chain initially. The trajectory of the MD simulation was used to generate starting structures for the non-adiabatic CT-QM/MM MD simulations. The time step of $\Delta t = 0.5$ fs was used and the length of the simulation between 2 ps and 20 ps with respect to the applied couplings and corresponding travelling time of the charge. The mobilities were obtained using a fitted MSD of 100 trajectories.

8.2. Temperature dependent mobility using Ehrenfest propagation

To benchmark the Ehrenfest propagation the non-adiabatic CT-QM/MM MD simulations are artificially modified to reduce the electronic couplings between selected molecules, while the obtained value for the reorganisation energy is unaffected.
The transport regime in the OSC depends strongly on the nature of the material, therefore, liquid crystalline systems are more in the hopping region and solid crystals are

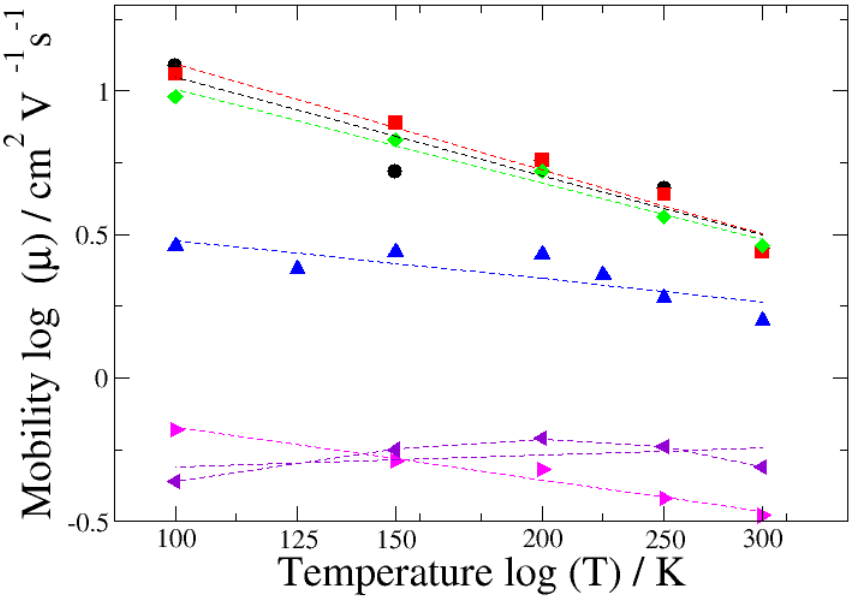

Figure 8.4.: Temperature dependent mobilities calculated by Ehrenfest method: calculated mobilities of the test system at different temperatures, presented as log(T) at x-axis and log(μ) at y-axis. The black dots correspond to the scaling of 1.54 of the coupling, red squares to 1.0, green diamonds to 0.75, blue triangles to 0.5 and purple and magenta rotated triangles to 0.25, respectively. The dashed lines indicate the linear fit. All simulations are performed with the multi-scale method applying Ehrenfest propagation.

between the hopping and band-like transfer. A pure band-like transfer happens in anorganic, mostly silicon based, semi-conductors. Therefore, the simulation method shall be able to treat the transfer region which lies between the both well-defined cases, hopping and band-like. As mentioned in the section 5, successful approach exist and can describe a purely hopping transfer and the same holds for a band-like transfers e.g. Holstein model.[204] Therefore, the goal is to develop a method which can handle both regimes and also the intermediate case.

The different scaling factors for the coupling, used to change the mobility of the charge carrier in the test system, are shown in Table 8.1. The simulations reproduced comparable values for the physically correctly scaled couplings, which were already obtained in Ref. 130,136 and are also shown in Table 8.1. The calculation of μ requires to calculate the diffusion constant D and therefore the MSD. The slope of the averaged MSD shows a linear regime which is used to determine the mobility with respect to Eq. 8.1. The MSD are presented in Figure 10.6 in the Appendix and the corresponding mobilities are visualised in Figure 8.4 and listed in Table 8.1.

Furthermore, the transport regime changes with respect to the scaled coupling. The simulations in this benchmark which has a coupling of 5.81$\pm$3.84 meV (scaling factor

Table 8.1.: Temperature dependent mobilities calculated by Ehrenfest method: the scaling factors, corresponding temperatures and resulting mobilities are listed. Each estimated slope is generated by an averaged MSD fitted to the linear regime.

scaling factor	Coupling / meV	Temperature	MSD slope / $\mathrm{Bohr}^2\mathrm{fs}^{-1}$	Mobility μ $\mathrm{cm}^2\mathrm{V}^{-1}$ s $^{-1}$
1.54	36.68±22.19	300	5.26	2.85 (2.93 Exp.)
		250	7.05	4.58 (3.74 Exp.)
		200	7.06	5.74 (5.05 Exp.)
		150	4.87	5.27
1.00	23.82 ±14.43	300	5.11	2.77
		250	6.79	4.41
		200	7.15	5.81
		150	7.17	7.76
		100	7.02	11.41
0.75	17.76 ±11.47	300	5.34	2.89
		250	5.64	3.66
		200	6.47	5.26
		150	6.21	6.73
		100	5.85	9.51
0.5	11.59 ±7.68	300	2.90	1.57
		250	2.90	1.88
		225	3.18	2.29
		200	3.32	2.70
		125	1.85	2.41
		100	1.76	2.86
0.25	5.81 ±3.84 (fit A)	300	0.90	0.49
		250	0.89	0.58
		200	0.76	0.61
		150	0.52	0.57
		100	0.27	0.44
0.25	5.81±3.84 (fit B)	300	0.61	0.33
		250	0.59	0.38
		200	0.59	0.48
		150	0.47	0.51
		100	0.41	0.66

of 0.25) show a much lower transport property. The trend of the monotonous decrease of the mobility with increasing temperatures disappears. Two different fits of the MSD are observed and mentioned as fit A and B. An initial similar slope is present for all simulations and after around 3 ps the higher temperature simulations offer another linear regime of the MSD. Therefore, using the first slope of the first 3 ps corresponds to a mobility between 0.66-0.33 $cm^2V^{-1}s^{-1}$ with increasing temperature, while using the last 2 ps the mobility is between 0.44-0.58 $cm^2V^{-1}s^{-1}$ with increasing temperature.

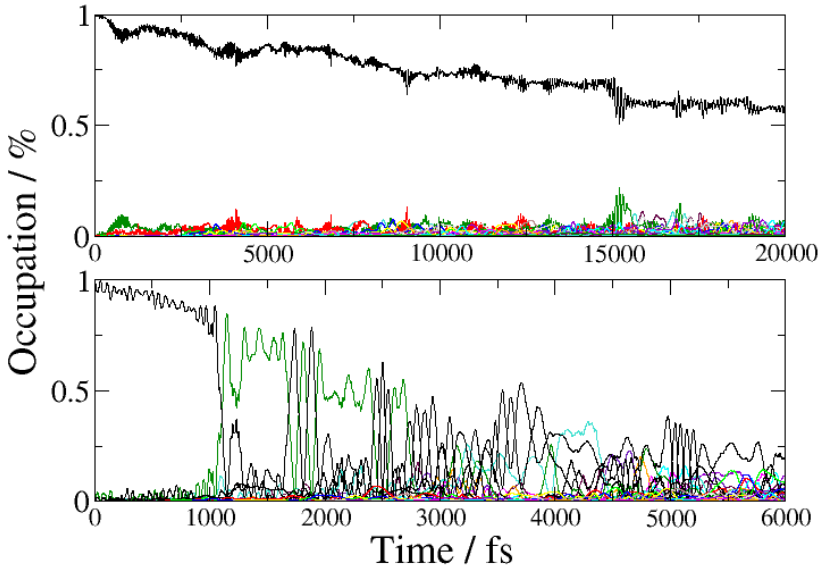

Figure 8.5.: Time evolution of occupation of the initially charge populated molecule site for one representative trajectory: the upper graph shows the constant reduced population of the initially populated site using the lowest coupling established by the scaling factor 0.1. The lower graph shows a randomly selected simulation of the 100 trajectories and the time-dependent change of population of the individual 30 sites, using the scaling of 0.25. The black line represents the charge population of the first site and the green one the second site. The x-axis presents is the evolution in time.

This rises the guess that a hopping transfer is achieved and longer simulations might be required. However, a deeper look in the individual simulations shows that the hopping character cannot be reproduced correctly. Figure 8.5 shows the occupation of the 30 sites in a randomly selected simulation, having the scaling factor of 0.25. The occupation of the first is reduced and the charge is transferred to the neighbouring site, which looks like a hopping transfer at 1 ps, however, the charge is not localised again at the neighbouring site. The charge is directly transferred to the next neighbouring site and this achieves a delocalisation over several sites soon. Thus, no localised hopping is obtained by this propagation method.
Additionally, 10 simulation with a lower scaled coupling of 0.1 are performed at 300 K with a propagation time of 20 ps, which corresponds to a coupling of 2.33 ± 1.55 meV.

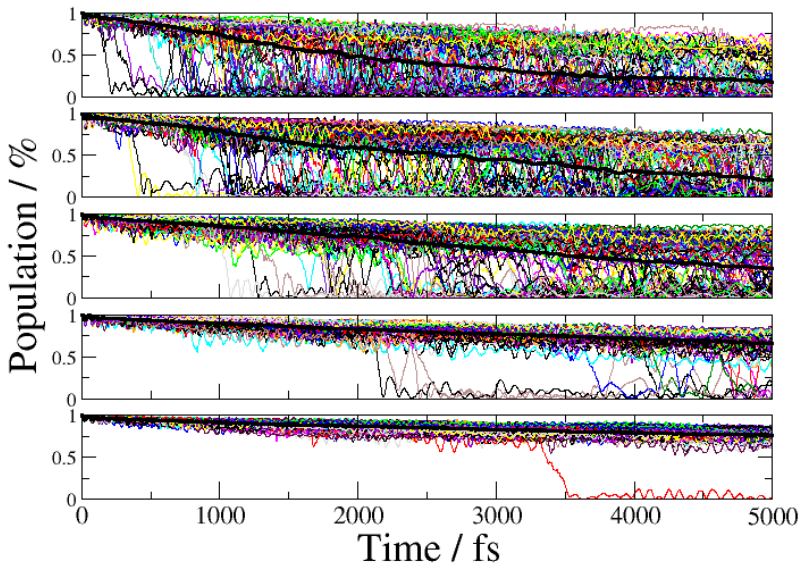

Figure 8.6.: Population of the first site of the Ehrenfest simulations (Scaling 0.25): the five plots correspond to different temperatures (first graph corresponds to 300 K, second to 250 K, third to 200 K, fourth to 150 K, and fifth to 100 K). The averaged population of the first site is presented by the thick, solid, black line in each plot. The coloured lines in the plots correspond to the reduced occupation of the first site of one (initially populated site) of the 100 simulations.

One of these simulations is also presented in Figure 8.5 and no transfer to the neighbouring sites is observed. A constant reduction of the occupation of the initially occupied site is visible, therefore, longer simulations might be required. It can be expected that the behaviours is similar compared to the 0.25 scaled simulations. Thus, the charge will spread soon after the first transfer, which will just take more time but not present a hopping-like behaviour, and therefore no further simulations are performed.
After the first transfer in the low coupling simulations, the charge does not relocalise again and spreads over the remaining system. Hence, this propagation cannot yield a hopping character in the organic crystalline system, because no included force does lead to the required relocalisation.

Nevertheless, the time until the first transfer happens, shows a temperature dependency, which is visualised in Figure 8.6. Therefore, the temperature dependent reduction of the initially occupied site can be used to approximately understand the charge transport mechanism in the test system and probably in further OCSs. Therefore, the population of the initially occupied site is averaged along all simulations and an exponential fit is applied. The slope of this fit can be used to calculate a rate for the first transfer. Marcus theory is widely used to determine the hopping rates, which is explained in section 5.3 and Eq. 8.5. The averaged values for λ, the coupling H_{MN} the temperature, T, and the Bolzman constant k_B are considered to calculate the rate con-

stants, k_{Marcus}, and compared to the rates obtained by the fit (all values are listed in Table 8.2)

$$k_{\text{Marcus}} = H_{MN}^2 \sqrt{\frac{\pi}{\lambda k_B T}}\, e^{\left(-\frac{\lambda}{4k_B T}\right)} \quad . \tag{8.5}$$

Table 8.2.: Temperature dependent population decrease of initially charged site. the scaling factor, corresponding temperatures, estimated fitted values and the calculated k_{Marcus} using Eq. 8.5 in ps are listed. The scaling of 0.25 corresponds to H_{NM}=5.81 meV, scaling of 0.1 to H_{NM}=2.33 meV and $\lambda = 84$ meV

Scaling	Temperature / K	Fit	k_{Marcus}
0.25	100	17.03	20.64
	150	9.26	13.31
	200	7.12	5.12
	250	6.24	3.38
	300	5.81	2.67
0.1	300	37.13	36.31

Overall, the obtained fit of the averaged reduced occupation of the initially populated site can be used to reproduce k_{Marcus}. Therefore, the assumption that the transfer happens between a dimer is assumed. The reduced occupation of the initially populated site is used but the increase of the other sites is not considered individually. In a homogeneous apolar environment the hopp between two molecules might be less affected by reorientations of the surrounding sites, and therefore the hopps might be similar, which supports this assumption. This shows that in this regime the averaged reduction of the initial occupied site can be used to estimate the rate for a hop to the neighbouring site with respect to the temperature.

8.3. Temperature dependent mobility using surface hopping propagation

The same test system, as in the benchmark of the previous section, is used to study CT process with another propagation method. The implemented SFH propagation achieves a more localised charge, whereas it is much more delocalised in the Ehrenfest approach. The scaled coupling also directly influences the mobility, μ. Furthermore, the estimated mobility is constantly underestimated by a factor of 5-6 compared to the experimental measured data, which is listed in Table 8.3 and illustrated in Figure 8.7. The charge stays localised over 2-4 sites during all simulations and does not spread over the hole system. Figure 8.7 shows the localisation of the charge carrier in a randomly selected SFH simulation, using the upscaled couplings and a temperature of 300 K, which

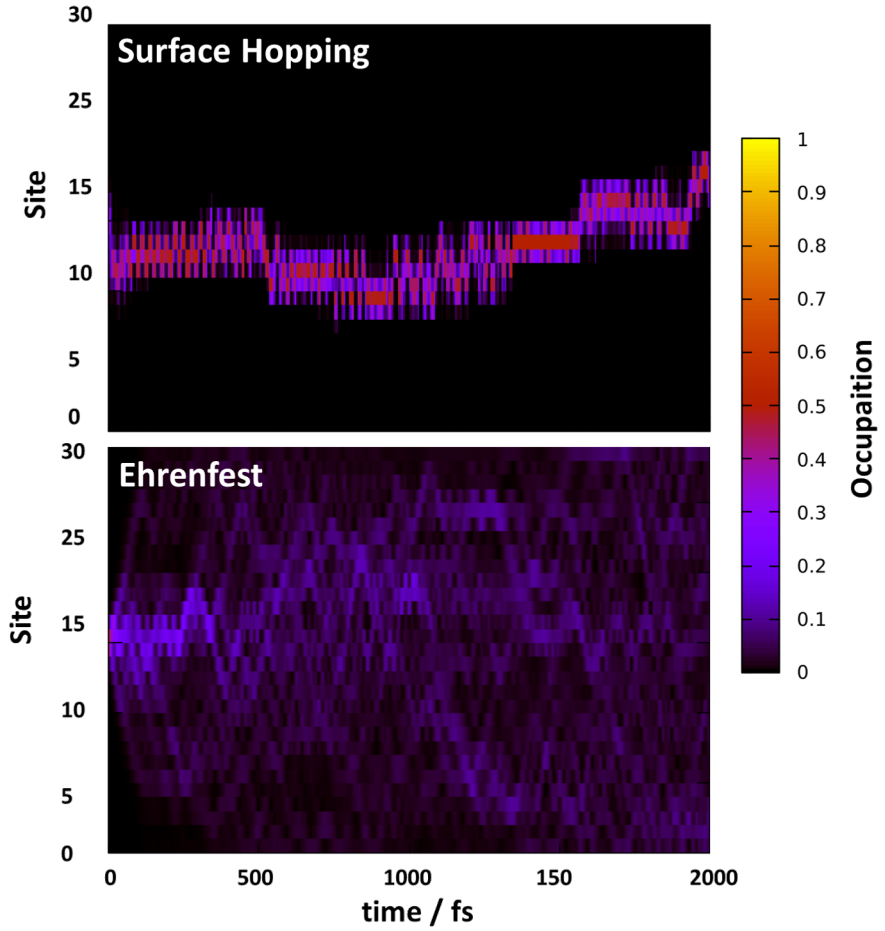

Figure 8.7.: Charge localisation during SFH and Ehrenfest propagation: the top part of the Figure presents the travelling charge in a randomly selected SFH simulation with upscaled couplings at 300 K. The lower Figure presents the Ehrenfest simulation using the same initial conditions than the SFH simulations. The colour code on the right side corresponds to the localisation and the y-axis on the left shows the position of the charge with respect to the molecules in the charge transfer pathway (30). The x-axis detonates the evolution of the simulation time.

directly presents the permanently relocalisation of the charge after a transfer between neighbouring sites.

Therefore, the simulation can be extended to longer timescales to study materials having a slower transfer. Furthermore, the correlation between hopping and a band-like transfer with respect to the decreasing couplings is also observed and visualised in the same manner as for the Ehrenfest propagation, shown in Figure 8.8. Therefore, the results of the 0.25 scaling are not shown which just show a further decrease in the mobility and a similar behaviour than the 0.5 scaled simulations.

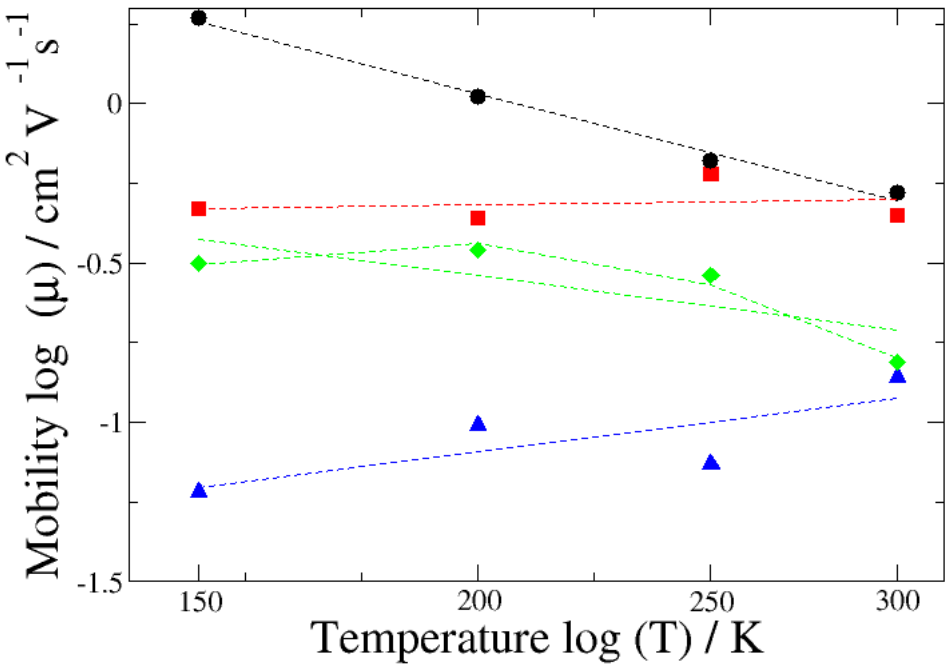

Figure 8.8.: Temperature dependent mobilities calculated by SFH method: calculated mobilities of the test system at different temperatures in b-direction, presented as log(T) at x-axis and log(μ) at y-axis. The black dots correspond to the scaling of 1.54 of the coupling, red squares to 1.0, green diamonds to 0.75, and blue triangles to 0.5, respectively. The dashed lines indicate the linear or exponential fit. All simulations are performed with the multi-scale method applying the flexible SFH propagation.

Table 8.3.: Temperature dependent mobilities calculated by SFH method: The scaling factors, corresponding temperatures and resulting mobilities are listed. Each estimated slope is generated by an averaged MSD fitted to the linear regime.

scaling factor	Coupling / meV	Temperature	MSD slope / $\mathrm{Bohr}^2\mathrm{fs}^{-1}$	Mobility μ $\mathrm{cm}^2\mathrm{V}^{-1}$ s $^{-1}$
1.54	36.68±22.19	300	0.98	0.53 (2.93 Exp.)
		250	1.02	0.66 (3.74 Exp.)
		200	1.29	1.05 (5.05 Exp.)
		150	1.74	1.88
		100	0.82	1.33
1.0	23.82 ±14.43	300	0.82	0.44
		250	0.92	0.60
		200	0.53	0.43
		150	0.43	0.47
		100	0.28	0.46
0.75	17.76 ±11.47	300	0.29	0.16
		250	0.44	0.29
		200	0.43	0.35
		150	0.29	0.31
		100	0.17	0.27
0.5	11.59 ±7.68	300	0.13	0.07
		250	0.06	0.04
		200	0.06	0.05
		150	0.03	0.03

8.4. Conclusions & Discussion

In the present study a test system was created and simulated with the non-adiabatic CT-QM/MM MD multi-scale method. Additionally, the method is modified and a surface hopping propagation (SFH) scheme is implemented. The available Ehrenfest and SFH propagation are benchmarked using the simulations of the test system. The couplings are reduced, while the reorganisation energy was constantly, just varying with respect to the molecular geometry ($\lambda \approx 83.6 \pm 4.5$ meV). The initial coupling H_{MN} is 36.68±22.19 meV which is scaled by a factor to differ the ratio with the reorganisation energy. Therefore, the benchmark of the propagation uses the test system to reproduce the dependency of the mobility, μ, with respect to the temperature and electronic coupling, H_{MN} and reorganisation energy λ.

The reduction of the coupling shows a sizeable effect in the transport of the nature of the charge mobility. The applied Ehrenfest propagation reproduced the experimental measured μ correctly, which was already shown for anthracene and further materials.[136] Furthermore, the determined mobility of the test system also decays with lowered couplings. However, the transfer to the hopping regime of the CT is not directly achieved. The missing hopping character is based on a well-known artificial over-delocalisation of the used Ehrenfest propagation. In the previous study of this problem was already discussed: it was purposed that an unphysical behaviour does not arise in the first 100-

500 fs simulation. Therefore, the mobility obtained by a linear fit of the MSD during 100-500 fs should be reliable. However, the spread of the charge over several residues can be seen as a polaron formation. Therefore, using just short simulation in which the delocalisation driven by the propagation might not induce an arbitrary behaviour and can yield the presented qualitative good results.[205,206] Especially in systems in which the CT might occur as a travelling polaron, additionally, this artificial spread could also overcome other errors and support the good resulting mobility. Studies of Prezhdo successfully introduced algorithms which introduce a decoherence correction into an Ehrenfest propagation scheme. This establishes a relocalisation of the charge carrier, which could be used in a further extension of this methodology to improve the accuracy further.[207]
On the first look the slope of Figure 8.4 underlays a hopping character, but a direct analysis of the occupation shows a different transfer regime. The charge is not relocalised after a transfer but directly transferred to more neighbouring sites. This behaviours is expected for Ehrenfest simulations in homogeneous environment. Nevertheless, an exponential fit of the averaged reduction from the population of the initially occupied site can be used to reproduced the hopping rates from the Marcus-equation. Thus, this propagation cannot be used to simulate hopping but it can be used to determine the rates for hopping with respect to the material properties (H_{MN} and λ).

The implemented Surface hopping propagation allows a simulation of a localised travelling charge, which spread between 2-4 molecules. This is an improvement of the Ehrenfest propagation. Thus, the charge is localised after a hop at the next sites before the next transfer happens. However, the obtained mobilities underestimate the measured mobilities by a factor of 5-6 and Ehrenfest slightly overestimates them, summarised in Table 8.4. The underestimation of the mobility could be also based on the too localised presentation of the charge carrier. Therefore, an optimisation of the obtained mobility could be achieved by modifying the used parameters in the propagation scheme to decrease the underestimation. This is not further investigated in this study because the focus was based on the observed transfer between the transport regimes.

Table 8.4.: The mobility μ / $\mathrm{Bohr}^2\mathrm{fs}^{-1}$ of the Ehrenfest and SH propagation with the upscaled coupling (1.54) are listed and compared to the experimental values.

Temperature / K	mobility μ / $\mathrm{cm}^2\mathrm{V}^{-1}$ s $^{-1}$ Experimental[208]	Ehrenfest	SFH
300	2.93	2.85	0.53
250	3.74	4.58	0.66
200	5.05	5.74	1.05

9. Proton transfer simulations in *E. coli* class-I CPD-Photolyase

Reproduced in part from Ref. 209

What accounts for the different functions in Photolyases and Cryptochromes: a computational study of proton transfers to FAD

with the permission of the Chem. Sci. Owner Societies.

The activation of Photolyases (PL) and Cryptochromes (CRY) is triggered by an electron transfer (ET) process, which partially or fully reduces the photo-activated FAD cofactor. The full reduction requires an additional protonation of the isoalloxazine ring of FAD. In plant CRY, an efficient proton transfer (PT) occurs, enabled by a conserved Asp working as proton donor. Spectroscopic studies of a plant CRYI from *Arabidopsis thaliana*, abbreviated as CRYI in the following, completely resolved the photo-induced ET and PT transfers: the photo-reduction of FAD_{ox} occurs within 31 ps and the subsequent PT from Asp396 to $FAD^{\bullet -}$ occurs within 1.7 μs, while the backward transfer is more than 100-times slower, happening within 690 μs.[61,74]

On the contrary, in class-I CPD-PL of *E. coli* PL, called PL-WT in the following, a PT happens at a second timescale without any obvious proton donor. Since the enclosed Asn cannot protonate FAD directly, a long-range PT pathway has to exist or can be formed transiently. Müller et al. investigated the PT and ET reactions that can occur in PL-WT *in vitro*, and reported detailed reaction times for all processes, beginning with an initially present FAD_{ox} and ending with a stabilised $FADH^{\bullet}$: the timescale of the initial photo-induced ET to form $FAD^{\bullet -}$ was below 300 ps and the additional subsequent PT happens in around 4 s. Thus, they happen clearly on different time scales and a coupled ET and PT can be excluded.[43] Afterwards, no deptotonation of $FADH^{\bullet}$ is observed and it is stabilised in the FAD pocket, which corresponds to the expected case since PL requires $FADH^{\bullet}$ to fulfil for the DNA repair.

Further experimental studies indicated that no other acidic residues, located in the FAD pocket, are directly involved in the PT.[58] Additionally, Damiani et al. suggested that

water molecules can enter the FAD pocket and form a PT pathway from the bulk water to FAD.[210]
A mutation in *E. coli* PL replaces the enclosed Asn with Asp, called PL-N378D in the following, and creates a FAD pocket which offers an additional possible proton donor enclosed to FAD. This mutation might form a similar pocket compared to the one of CRYI. Experimental studies of PL-N378D showed that the photo-induced ET proceeds similar to PL-WT, forming $FAD^{\bullet -}$ within the picoseconds regime. However, no PT and no $FADH^{\bullet}$ stabilisation happens afterwards. The formation of a FAD^{x} state occurs in 3.3 μs, which surprisingly shows that the inserted Asp does not decrease the transfer time of the protonation. Furthermore, this mutation cannot work as proton donor and also impedes the protonation of $FADH^{\bullet}$ along any other pathway completely and stabilises the FAD^{x} state. This highlights that the pocket of PL-N378D cannot form or stabilise $FADH^{\bullet}$ in the same fashion as the PL-WT or CRYI.

Thus, the nature of the FAD pocket facilitates different processes in the different systems which are schematically shown in Figure 9.1: in CRYI the switch between protonated and deprotonated FAD is achieved, in PL-WT a stabilisation of $FADH^{\bullet}$ is assured and in PL-N378D a structural rearrangement, shown by the presence of the FAD^{x} spectra is obtained, however, no stabilisation of $FADH^{\bullet}$.

Computational chemistry methods can provide a consistent and comparatively cheap approach, to explore this unexpected behaviour of the PL-N378D mutant and the alternative long-ranged PT mechanism in the PL-WT. Furthermore, a comparison to the direct PT in CRYI can be obtained. The combination of several computational methods can explain the divergent behaviour towards the FAD protonation: classical MD simulations allow deeper insight into the rearrangements after the photo-induced ET in the FAD pocket and classical biased MD simulations can be used to study the reaction times for the subsequently observed reorientations. Additionally, free energy landscapes of the PT events are obtained from state-of-the-art biased QM/MM simulations and the comparison of the computed free energy barriers with kinetic and spectroscopic measurements allows a microscopic interpretation of the occurring structures and processes.
Based on that, the results can be summed up to illustrate the fine-tuning of the electrostatic FAD environment and the adaptability of the FAD pocket to ensure the divergent functions of the members of the PL/CRY family.

In a similar study of Dr. G. Lüdemann[18] in CRYI the computed rate constants for both, ET[16–18] and PT,[18] were in agreement with the experimental results. The aim of this chapter is to study and resolve the mechanism and energetics of the PT processes taking place in PL-WT and PL-N378D. Additionally, the obtained results are compared to those with the studied CRYI (see Figure 9.1). Based on that comparison, the impact of the slight structural differences of these proteins revels a better understanding of the

resulting functionality. Therefore, the aim of this study is to solve the experimentally raised questions by Müller, explained in the following:[43]

- Why can PL-N378D not protonate $FAD^{\bullet -}$ while the related CRYI, having the same proton donor, can protonate $FAD^{\bullet -}$? And which property of the FAD pocket may hinder the PL in PL-N378D?
- How are the energetics of the PT in PL-N378D?
- Does the poto-induced ET trigger structural reorientations in PL-WT which can explain the long time scale for the PT?
- Which pathway is used for the protonation in PL-WT?

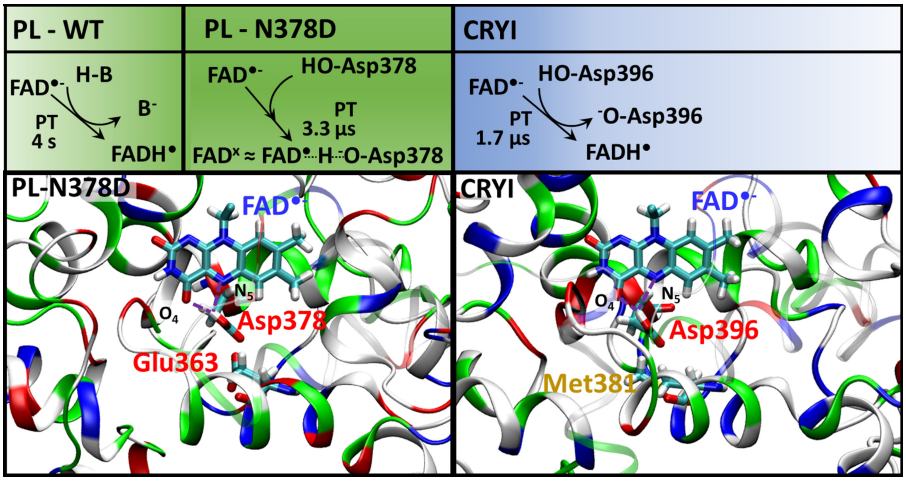

Figure 9.1.: Comparison of the $FAD^{\bullet -}$ protonation in *E. coli* PL-WT, PL-N378 mutant (green) and CRYI (blue): the nature of the proton donor is given wherever identified, otherwise B stands for a generic proton donor. The FAD binding pocket of PL-N378D and CRYI is shown with amino acids coloured white (non-polar), green (polar) and red/blue (acidic/basic). The isoalloxazine ring of FAD and the side chains of Asp378/396 and Glu363 or Met381 are displayed explicitly. Purple dashed lines indicate hydrogen bonds between N_5 of FAD and Asp378/396. All the proteins pictures were obtained using VMD.[145]

9.1. Computational details

9.1.1. Model structures and MD simulations

The starting structures for MD simulations was adapted from the X-ray crystal structures of *E. coli* class-I CPD-PL (PDB ID:1DNP)[59] and CRYI *Aarabidopsis thaliana* (PDB ID:1U3D)[211] by Deisenhofer and co-wokers. The protonation states of relevant amino-acid side chains were established on the basis of pKa calculations performed with PROPKA3.1[192,193] on protein structures taken from pilot MD simulations. Since no crystal structure of PL-N378D mutant of *E. coli* PL was available, a starting structure was created by modifying the PL-WT structure.

In *E. coli*, two different starting structures were created containing different protonation states of Glu363. Glu363 has a calculated pKa value of 6 which suggests the coexistence of the protonated and the deprotonated states. The different protonation states of Glu363 results in different interaction patterns with the carboxyl group of Asp378 and the surrounding solvent: the protonated Glu363 is able to form a strong hydrogen bond with negatively charged Asp378, and a weaker interaction with neutral Asp378; a deprotonated Glu363, however, may lead to a destabilisation of the deprotonated Asp378 by electrostatic repulsion. Furthermore, another mutation was obtained in the PL-N378D substituting the Glu363 into a methionine, Met363, which is the corresponding amino acid in the CRYI.

All MD simulations were performed with the AMBER-SB99-ILDN force field[149,150,190] using GROMACS 5.0.4.[147,148] The cofactors FAD and MTHF were parametrised with the *xLeap* module of AmberTools[152,212] employing the general Amber force field (GAFF).[151,152] The atomic charges of FAD, $FAD^{\bullet-}$ and $FADH^{\bullet}$ were fitted on the electrostatic potential (RESP)[153,191] obtained at the HF/6-31G*[154,165] level of theory with Gaussian09.[155]

The protein was placed in a periodic box sized 96x94x121 $Å^3$ for PL-WT and PL-N378D and 98x98x98 $Å^3$ for CRYI. The box was filled with TIP3P water molecules suitable to obtain a density of 1000 $kg \cdot m^{-3}$, and an appropriate number of water molecules were substituted by sodium ions to achieve electroneutrality; no extra salt was added. These systems were equilibrated by means of a protocol consisting of a series of energy minimisations, NVT (1 ns) and NPT (1 ns) simulations. The eventual, production simulation uses the Nosé-Hoover thermostat[112] and the Parrinello-Rahman barostat[113] to maintain a temperature of 300 K and a pressure of 1 bar, respectively. The simulations employed a leap-frog integrator with a time step of 2 fs. Initial MD simulations (100 ns) for the PL-WT, PL-N378D and CRYI contained FAD_{ox}. Then, the last geometry from these simulations were used to start free MD simulations for CRYI and PL-N378D (200 ns) and free MD simulations for PL-WT (50 ns) in which the parameters were changed to $FAD^{\bullet-}$.

Coulombic energy calculations in PL-N378D and CRYI

The GROMACS tool gmx energy was used to evaluate the electrostatic interactions between the groups of atoms defined as follows: side chain of Asp (for PL-N378D Asp378 and for CRYI Asp396), isoalloxazine ring of FAD, solvent and protein (excluding the side chain of Asp). The electrostatic interactions between each pair of the groups were calculated using a cut-of radius of 10 Å along the last 20 ns of the MD trajectories for each protein for the $FAD^{\bullet-}$ state. The 20 ns trajectories were used to generate snapshots each 2 ns, which served as starting structures for individual MD simulations, containing $FADH^{\bullet}$ and deprotonated Asp378/394. This generates 10 individual MD

simulations and each contained the force field parameters of $FADH^{\bullet}$ and deprotonated Asp378/394. Then, they were used to create a trajectory of an accumulated length of 20 ns.

9.1.2. Biased sampling methods

Free energies of the examined processes, structural rearrangements and PT reactions, were obtained with umbrella sampling (US) or well-tempered (WT) metadynamics,[115,117] whichever proved more appropriate in pilot simulations. The free energy was obtained directly in metadynamics simulations, while it was estimated by using the weighted histogram analysis method (WHAM)[213,214] in US simulations. All of the biased sampling simulations were performed with GROMACS interfaced with the Plumed 2.0.1 software.[215,216]

Classical WT metadynamics simulations of the N378 side chain rotation in PL-WT

Classical WT metadynamics simulations (50 ns) were performed to obtain the free energy profile of the Asn378 side chain rotation. The dihedral angle $\chi = \mathrm{N}_{\delta}-\mathrm{C}_{\gamma}-\mathrm{C}_{\beta}-\mathrm{C}_{\alpha}$ was used as reaction coordinate. Gaussian shaped hills were added every 500 ps, with a width of $\sigma = 20^{\circ}$, an initial height of 0.29 kcal·mol^{-1}, and the bias factor was set to 6.

Classical umbrella sampling simulations of a water wire formation in PL-WT

Classical US simulations were used to generate the free energy profiles of the penetration of water molecules into the FAD binding pocket. The reaction coordinate for the entrance of the first water molecule was the distance between the centre of mass of a selected water in front of the loop, being 6.5 Å away, and the centre of mass of the N_5 and O_4 atoms of FAD. The interval of 6.5 Å was divided into 29 windows. Harmonic biasing potentials with a force constant of 23,9 kcal·mol^{-1}·Å^{-2} were applied, and each window was simulated for 50 ns. The entrance of the second water was also studied with the same US approach and the reaction coordinate was the distance between the firstly inserted water and the second water. The first water was fixed inside the pocket 3 Å away from FAD N_5 using a spherical position restrain.

9.1.3. QM/MM simulations

PT reactions were simulated with the quantum chemistry molecular mechanics approach (QM/MM) employing the approximative DFT method for the QM part, the third-order density functional tight-binding (DFTB3) approximation.[99] The general parametri-

sation, 3OB, of DFTB3 for organic and biomolecules which exhibits a sufficient accuracy for biophysical application was used.[107] In general, DFTB underestimates the barrier for PT because its parametrisation is based on a GGA PBE functional.[104] Nevertheless, DFTB3/3OB provides a good description of proton affinities, which were relevant for the simulation of PT reactions, as well as PT barriers. All QM/MM simulations were performed with a QM/MM implementation of DFTB3 in GROMACS combined with Plumed.[188] Hence, it was possible to perform US and metadynamics simulations with a QM/MM Hamiltonian (see section 4.2 for further details of the methodology).

QM/MM umbrella sampling simulations of proton transfer

In PL-WT, the free energy profile of the PT reaction along the water wire in the protein pocket was obtained with an US-QM/MM simulation. The QM region consists of the isoalloxazine ring of FAD, the side chain of Glu106 and five water molecules, and it was described with DFTB3/3OB. The progress of the PT process is captured with the centre of excess charge reaction coordinate ζ, see section 9.4.3.[217] The distance between proton acceptor and donor was divided into 57 windows. Harmonic biasing potentials with a force constant of 239 $\mathrm{kcal \cdot mol^{-1} \cdot Å^{-2}}$ were applied, and each simulation was extended to 300 ps.

In PL-N378D, the free energy profile of the PT reaction between N_5 of the FAD and the Asp378 was obtained with a US-QM/MM simulation. The QM region contains the side chain of Asp378 and the isoalloxazine core of FAD. The difference between the H-O distance of Asp378 and the H-N_5 distance of $FAD^{\bullet -}$, $d_{\mathrm{O\text{-}H\text{-}N}}$, was considered as a reaction coordinates for the PT (anti-symmetric stretch coordinate). These simulations were performed for the three protein variants involving a protonated Glu363, a deprotonated Glu363 and an inserted Met363 instead of the Glu363. The replacement of Glu363 into Met363 converts the pocket of the PL even more into the pocket of the CRYI. Therefore, this double mutant can have an sizeable influence of the PT properties. Harmonic biasing potentials with a force constant of 239 $\mathrm{kcal \cdot mol^{-1} \cdot Å^{-2}}$ were applied, and the simulations were extended to 300 ps for each of the 20 windows. The shown data of the PT in CRYI was based on the previous study,[18] in which only the distance between the proton and the N_5 of FAD, $d_{\mathrm{H\text{-}N}}$, was considered as reaction coordinate.

9.2. Structural analysis of the FAD pocket in PL-N378D and CRYI

The experimental time scale for the protonation of $FAD^{\bullet -}$ is 1.7 μs in CRYI and the formation of FAD^{x} in PL-N378D occurs within 3.3 μs Thus, the protein is able to relax after the photo-reduction of FAD. In a first step, unbiased classical MD simulations are used

to equilibrated the CRYI and *E. coli* mutant (PL-N378D) structures with neutral FAD_{ox}. In the next step, to model the $FAD^{\bullet -}$ states, unbiased MD simulations from the respective neutral structures are performed, considering a non protonated semi-quinone $FAD^{\bullet -}$ in the active site. This generates the starting point for the PT reaction, in which the Asp378/396 backbone carbonyl oxygen faces the neutral FAD.
After the photoinduced ET, forming a negatively charged $FAD^{\bullet -}$, the Asp residue in both proteins rotates within few nanoseconds. Hence, the protonated Asp side chain now faces the O_4 or N_5 of FAD and forms a hydrogen bond, which is visualised in Figure 9.2. This structural reorientation is prerequisite for a potential PT forming $FADH^{\bullet}$,

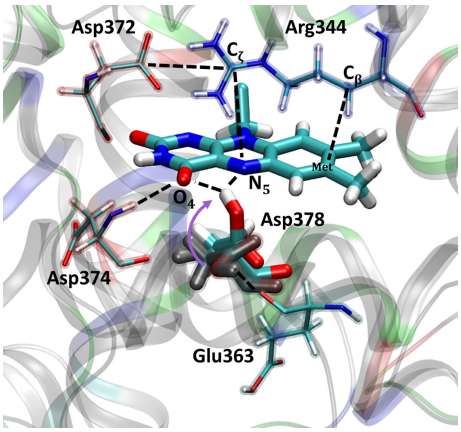

Figure 9.2.: FAD binding pocket of PL-N378D: the grey coloured position of Asp378 belongs to the neutral orientation and the purple arrow indicates the rotation. Dashed lines indicate distances which are calculated during the simulation and listed in Table 9.1. The numbering of Asp374, Asp372, Asp378 and Arg344 corresponds to PL-N378D, while respectively in CRYI Asp392, Asp390, Asp396 and Arg362. The position of Glu363 in PL-N378D corresponds to the Met381 in CRYI.

which was already observed for CRYI[18] and is now also seen in the PL-N378D mutant. Therefore, the surrounding of Asn378 in PL does not impede this rotation, which derives the assumption that another source prevents the formation of $FADH^{\bullet}$.

The distance between FAD N_5 and the proton of Asp378/396 is calculated along the MD trajectories of both systems, PL-N378D and CRYI. Additionally, the distances between FAD and several neighbouring amino acid side chains are calculated and the averaged values are listed in Table 9.1. In CRYI, the distance between the proton of Asp378/396 and N_5 is on average smaller than in PL-N378D, where on the contrary the distance between the proton and O_4 of FAD is smaller. Indeed, the proton of Asp396 in CRYI faces more often FAD N_5 than in PL-N378D.
The neighbouring amino acids (Arg344-Asp372 in PL-N378D and Arg362-Asp390 in CRYI) show a conserved salt bridge. The relative positions of Arg344/362 and FAD differs in the two proteins. Comparing the crystal structure of CRYI and the PL-WT (PL-N378D is not crystallised): the distance between C_β of Arg344/362 and the methyl group of the FAD isoalloxazine ring, FAD_{Met}, is 0.5 Å larger in CRYI than in PL-N378D. During the MD simulation, the distance increases to more than 1 Å. Additionally, the

Table 9.1.: Averaged distances and corresponding standard deviations along the 200 ns trajectory including the $FAD^{\bullet-}$ state after a 100 ns simulation with neutral FAD. The numbering of Asp374, Asp372 and Asp378 corresponds in CRYI to Asp392, Asp390 and Asp396.

			PL-N378D Distance / Å		CRYI Distance / Å	
FAD N5	-	Arg-C_ζ	4.42	$\pm$ 0.23	4.86	$\pm$ 0.77
FAD_{Met}	-	Arg-C_β	3.85	$\pm$ 0.21	5.08	$\pm$ 0.34
Arg-C_ζ	-	O-Asp372	3.98	$\pm$ 0.10	4.16	$\pm$ 0.30
FAD O4	-	H-Asp374	2.25	$\pm$ 0.28	3.00	$\pm$ 0.18
FAD O4	-	H-Asp378	1.91	$\pm$ 0.54	1.99	$\pm$ 0.46
FAD N5	-	H-Asp378	2.79	$\pm$ 0.52	2.64	$\pm$ 0.40

guanidine group of the Arg362 side chain is on average 0.5 Å farther away from FAD N_5 in CRYI compared to PL-N378D.
This difference in the geometry of the pocket can cause a different electrostatic interactions due to the positive charge of Arg344/362. Hence, a detailed analyse is performed and shown in the next section.

A further difference of the geometry in the pocket is exposed by the standard deviations of the distances between the proton of the Asp and the Arg-Asp salt bridge, which are larger in CRYI, indicating a higher structural flexibility. Additionally, earlier experimental studies of PL and CRY also described the FAD pocket in CRYI to be more flexible, which is in agreement with the calculations of the free MD Simulations.[218,219]

9.3. Proton transfer in PL-N378D

The ET along the Triad reduces the excited FAD_{ox} and induces the rotation of the O-H proton, described above. This brings Asp378 into a position where a PT to the N_5 of FAD is possible.
To study the PT-energetics, the one dimensional free energy landscape of the PT in PL-N378D is computed using biased QM/MM MD simulations (see section 9.1 for details). The results are presented in Figure 9.3 and compared to the previous results from CRYI, which were obtained by Dr. G. Lüdemann.[18]

The corresponding driving force ΔG^0 and activation free enthalpy $\Delta G^{\neq}$ values are listed in Table 9.2. The three states labelled I., II. and III. designate the reactants with the proton on the aspartic acid, the transition state (TS) and the products with the proton on $FADH^{\bullet}$, respectively. An important difference is clearly apparent in the Gibbs free energy of the PT: while state III. is the global minimum in CRYI, state I. is favoured in PL-N378D. In PL-N378D, Glu363 is considered protonated because it showed the smallest ΔG^0. Additionally, the free energy profile of deprotonated Glu363 was calculated but will not be taken into account in the next analysis because it does not support the PT and is

shown in the Appendix in Figure 10.8. The deprotonated Glu363 yields a higher barrier and smaller ΔG^0.

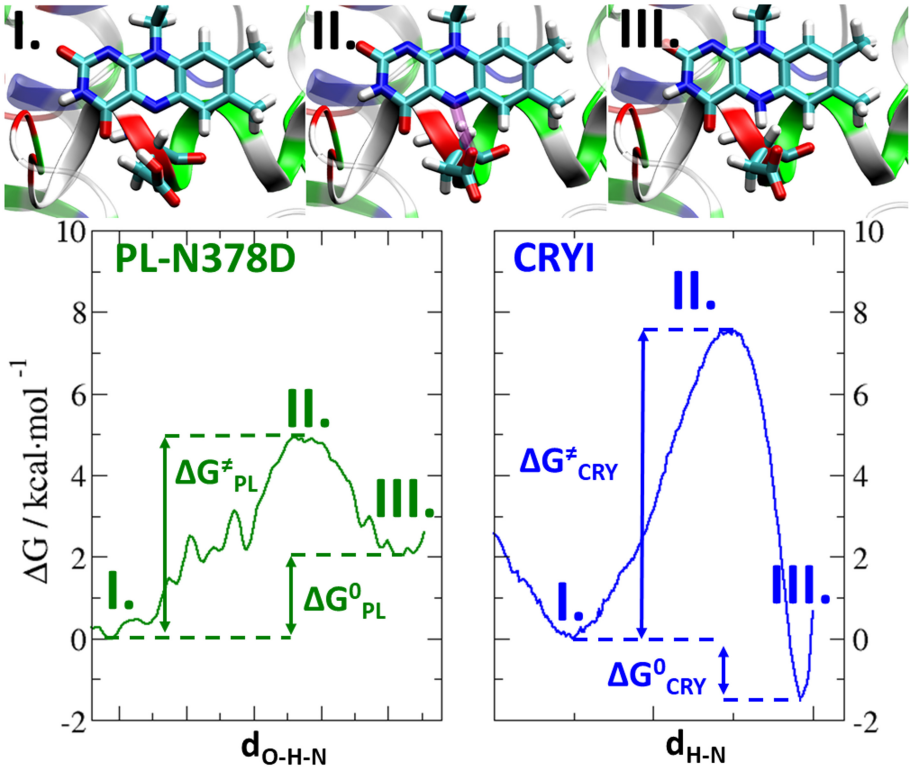

Figure 9.3.: Free energy landscape of proton transfer in PL-N378D and in CRYI: the different states are numbered: I. reactants, II. transition states, and III. products. The reaction energies, ΔG^0, and the barrier heights, $\Delta G^{\neq}$ are indicated by arrows.

The obtained energy barrier can be used to calculate the rate constant for the PT by applying the transition state theory, using the relation[220], which is applied as in the same fashion in Ref. 18:

$$k = \frac{k_b T}{h}(1 - \exp^{\frac{-h\nu}{k_b T}}) \exp^{\frac{-\Delta G^{\neq}}{k_b T}} \tag{9.1}$$

where T is the temperature, $\Delta G^{\neq}$ the barrier height, Boltzmann's constantsand k_b and Planck's constants h. The attempt frequency ν of the transition is estimated as

$$\nu = \frac{1}{2\pi}\frac{\sqrt{k_{H-N}}}{m_H} = 7.5 \cdot 10^{13} s^{-1} \tag{9.2}$$

with $k_{H-N} \approx 537\,\text{kcal·mol}^{-1}\,\text{Å}^{-2}$ being the harmonic force constant of the N-H bond in the Amber force field and m_H being the mass of a hydrogen atom.

PT is an endergonic reaction in PL-N378D but exothermic in CRYI. Therefore, the computed energy landscape allows an interpretation of the absorption spectra, measured in CRYI and PL-N378D. CRYI shows slow rate constants of $0.02\,\text{ns}^{-1}$ and $0.002\,\text{ns}^{-1}$ for the forward and backward PT respectively, and a well-defined $\text{FADH}^{\bullet}$ spectrum, which

is consistent with the free energy landscape favouring $FADH^{\bullet}$.

For PL-N378D, the reported measured spectrum was called $FAD^{x\,43}$, which strongly overlaps with the $FAD^{\bullet-}$ spectrum, however, showing only a slight deviation from the spectrum of a pure $FAD^{\bullet-}$ species. This can be interpreted with the free energy profile, highlighting an energetically favoured $FAD^{\bullet-}$ state. Nevertheless, the activation energy for the forward PT is just 4.9 kcal·mol^{-1} and only 2.9 kcal·mol^{-1} for the backward PT which corresponds to rate constants of 1.6 ns^{-1} and 47 ns^{-1} respectively. Therefore, PT is feasible during the recording times of the spectra, and a superposition of both species may lead to the particular form of the absorption spectra.

Table 9.2.: The free energies and rate constants calculated from the biased QM/MM simulations of PT in PL-N378D and in CRYI.

			PL-N378D	CRYI
$\Delta G^{\neq}_{\mathrm{I.}\rightarrow\mathrm{II.}}$	($G_{\mathrm{II.}}$-$G_{\mathrm{I.}}$)	/ kcal·mol^{-1}	4.9	7.5
$\Delta G^{\neq}_{\mathrm{III.}\rightarrow\mathrm{I.}}$	($G_{\mathrm{II.}}$-$G_{\mathrm{III.}}$)	/ kcal·mol^{-1}	2.9	8.9
ΔG^{0}	($G_{\mathrm{III.}}$-$G_{\mathrm{I.}}$)	/ kcal·mol^{-1}	2.0	-1.4
ΔG^{0}_{PL} - ΔG^{0}_{CRY}		/ kcal·mol^{-1}	3.4	
$k_{\mathrm{I.}\rightarrow\mathrm{III.}}$		/ ns^{-1}	1.6	0.020
$k_{\mathrm{III.}\rightarrow\mathrm{I.}}$		/ ns^{-1}	47.0	0.002

Since it was not possible to generate the CRYI functionality with a single mutation in PL, another mutation is created to further study the PT in PL-N378D. The protonation state of Glu363 influences the PT between Asp378 and $FAD^{\bullet-}$ in PL-N378D by lowering the barrier of up to 2 kcal·mol^{-1}. Thus, this residue could partially explain the different behaviour between both proteins.
While CRYI shows an apolar Met381 at the homologous position, a double mutant appears to be a good candidate to enable the PT ability of CRYI in the PL-N378D pocket. Therefore, Glu363 is mutated into Met, nevertheless, no improved stabilisation of the PT product is observed (free energy curve is shown in Figure 10.8 in the Appendix).

The mutation rebuilds the CRYI pocket out of the PL-WT by introducing a proton donor and replacing a polar amino acid into an apolar one which are two obversely largest differences around the PT active site. However, the ability of the pocket could not be observed. Taken together, this double mutation is also not sufficient to convert the nature of PL-WT into CRYI, therefore, a combination of multiple small contributions seems to be necessary to make the PT reaction feasible.

9.3.1. Analysis of Coulomb interactions between FAD and the pocket

The considered PT between $FAD^{\bullet-}$ and Asp378 involves a partial transfer of negative charge from the large isoalloxazine ring to a small carboxyl group. Therefore, the electrostatic interactions between the environment and the active site may play an im-

portant role in the stabilisation of the protonation states. To quantify these effects, the electrostatic interaction energies, E_{Coul}, between the PT complex (isoalloxazine ring and Asp378/396) and the protein and solvent are calculated for the FADH$^{\bullet}$ and the FAD$^{\bullet-}$ state.

Thus, the structures of the individual trajectories are used to calculate averaged E_{Coul} of each protonation state and are compared according to Eq. 9.3. The difference, ΔE_{Coul}, between the product and reactant state illustrates the contributions of the individual interactions which stabilise the product or the reactant.

$$\Delta E_{Coul} = E_{Coul}([FADH^{\bullet} - ASP^{-}]) - E_{Coul}([FAD^{-} - ASP - H]) \tag{9.3}$$

In CRYI, the FADH$^{\bullet}$–Asp378^{-} product is slightly favoured (the sum of ΔE_{Coul} = –3.9 kcal·mol^{-1}) and no significant preference for either state in PL-N378D is observed (sum of ΔE_{Coul} = +0.6 kcal·mol^{-1}). It is interesting to see that the total difference in electrostatic energy, ΔE_{Coul}, of 4.5 kcal·mol^{-1} roughly resembles the difference in the reaction energy as shown in Table 9.2. The analysis of the different contributions, listed in Table 9.3, shows that the major difference between both proteins is based on the interaction between Asp^{-} and the environment. Furthermore, the ΔE_{Coul} interactions between the

Table 9.3.: ΔE_{Coul} calculations between the selected groups: Asp-FAD-complex, the protein and the solvent for the different protonation states of FAD. The trajectories of the classical MD simulations are used for the calculations.

Coulomb energy kcal·mol^{-1}		FAD$^{\bullet-}$–ASP-H		FADH$^{\bullet}$–ASP^{-}			
		Educt		Product		ΔE_{Coul}	
System		CRYI	PL-N378D	CRYI	PL-N378D	CRYI	PL-N378D
FAD	Solvent	-85.10	-87.76	-52.38	-54.28	32.72	33.48
ASP	Solvent	0.26	-0.15	-55.81	-2.20	-56.07	-2.05
FAD	Protein	-19.98	-33.72	-19.54	-30.17	0.44	3.56
ASP	Protein	-107.89	-147.05	-115.06	-193.13	-7.17	-46.08
FAD	ASP	-64.93	-63.47	-51.22	-49.69	13.70	13.78
$\sum$		**-277.64**	**-332.15**	**-294.02**	**-329.47**	**-16.37**	**2.68**

protein and the complex are split to observe the individual interactions between each amino acid and the complex. Thus, the pocket of both proteins is presented in Figure 9.4 and especially the ΔE_{Coul} of the individual amino acids and water molecules are coloured in red or blue with respect to their positive or negative values. Additionally, the individual interactions between the amino acids and the complex is also visualised in Figure 9.5, in which the residue number is plotted against the ΔE_{Coul}.

No positive value of ΔE_{Coul} larger than +0.36 kcal·mol^{-1} is observed in CRYI, which could disfavour the FADH$^{\bullet}$ state. However, in PL-N378D several positive contribu-

tions occur (+0.3, +0.4 and +2.96 kcal·mol^{-1} for Leu375, Gly381 and Arg344, respectively). Thus, Arg344 (Arg362 in CRYI) interacts differently: in CRYI it favours PT(-1.746 kcal·mol^{-1}) and on the contrary it disfavours in PL-N378D. The distance between Arg and FAD, which was previously mentioned in Table 9.1, is shorter and the structure is more rigid in PL-N378D. This could explain part of the preference for the negatively charged FAD$^{\bullet-}$ state in CRYI.

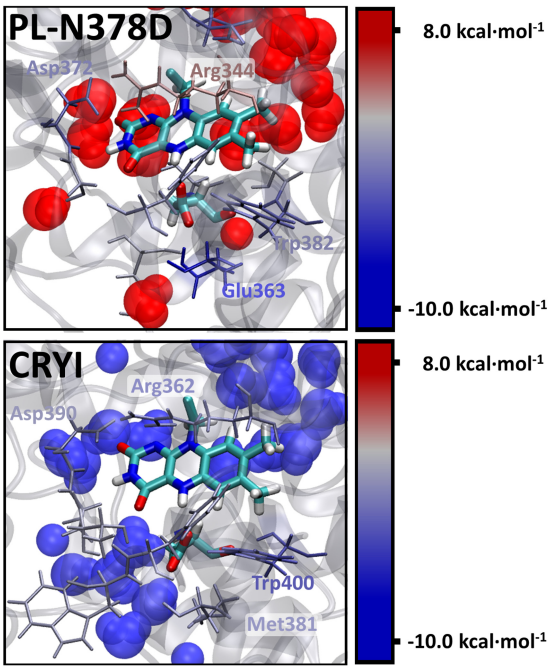

Figure 9.4.: Structural view of the FAD binding pocket in CRYI and PL–N378D: amino acids are coloured in blue, grey and red corresponding to negative, neutral and positive ΔE_{Coul}, respectively. The side chain of all residues with $|\Delta E_{Coul}| > 0.25$ kcal·mol^{-1} are represented in sticks. Close water molecules are also included and coloured regarding to the averaged interaction with FAD and Asp378/396 which is in CRYI -5.58 and +7.51 kcal·mol^{-1}in PL-N378D.

Negative values of ΔE_{Coul}, favouring the FADH$^{\bullet}$ formation, are observed similarly, such as the interaction with the first Trp of the Triad. Furthermore, in PL-N378D the Glu363 ($-10.32\,\mathrm{kcal\cdot mol^{-1}}$) is strongly hydrogen bonded to the negatively charged Asp378, while in CRYI Met381 ($1.05\,\mathrm{kcal\cdot mol^{-1}}$) just slightly interacts with the FAD-Asp complex. This stabilising interaction relies on the protonation state of Glu363, while assuming a protonated Glu363, however, a deprotonated Glu363 would disfavour the PT and therefore stabilise the FAD$^{\bullet-}$ state more.
In CRYI, the Met381 may seems not to be a good choice for stabilisation of the FADH$^{\bullet}$. However, Met381 shows a flexibility allowing water molecules to approach Asp396, which results in a stabilising interaction of -5.58 kcal·mol^{-1} with negatively charged Asp. In contrast, the interaction with water molecules rises a positive E_{Coul} (+7.51 kcal·mol^{-1}) in the PL-N378D, destabilising the product state (averaged values of Table 9.3).

Taken together, the analysis shows three major factors being responsible for the difference in the E_{Coul}: the different orientation of the Arg as mentioned above, the interaction

between Asp378 and Glu363 and a different interaction pattern with water molecules in the pocket is possible in CRYI.

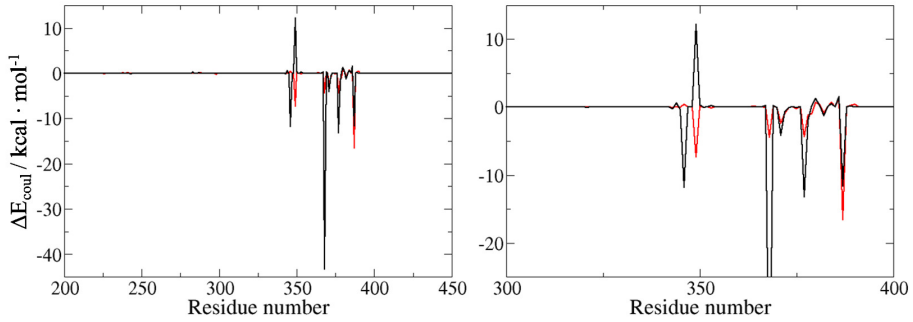

Figure 9.5.: Electrostatic interaction, ΔE_{Coul}, between each side chain of the protein and the PT complex (FAD and Asp378/396): the left part shows the interaction between all amino acids and the right a zoomed in view of the important, stronger interacting residues (300-400) in the pocket.

9.4. Proton transfer in PL-WT

The mentioned observations in the PL-N378D mutant and in CRYI,[18] which are triggered by the ET forming $FAD^{\bullet-}$, support the hypothesis that this ET may also induce structural reorientations in the PL-WT. In addition to that, the experimental observed timescale of seconds[43] and the fact that no enclosed proton donor is available, points out that the PT cannot happen without a rearrangement or transient pattern formation of a pathway.

In CRYI and the PL-N378D, a rotation of Asp378/396 occurs within a few nanoseconds, triggered by the negative charge of FAD. In PL-WT, Asn378 takes the respective position and is hydrogen bonded to the FAD N_5. Therefore, 10 independent 50 ns classical MD simulations are performed to assess the relaxation of the protein structure in response to a photo-induced ET forming $FAD^{\bullet-}$. The starting structures are taken from a MD trajectory containing FAD_{ox} and the switch into the charge state $FAD^{\bullet-}$ happens by changing the corresponding set of MM atomic charges.

In the starting structure of the MD simulations the amido oxygen of Asn378 points towards N_5 of $FAD^{\bullet-}$, which is shown in Figure 9.7 and will be call O-conformation in the following. In 9 out of 10 simulations, significant structural rearrangements are observed in the area enclosed by FAD, Asn378 and a loop close to the heterocycle of the isoalloxazine ring. The following two different events occur, one directly influences the protein structure in the pocket and the other is more driven by the surrounding water:

- Rotation of the Asn378 side chain to face the FAD N_5: in the initial state, the keto group faces N_5 and after the rotation the side chain amino group of Asn is facing towards N_5 (7 occurrences). This rotation, which will be further mentioned as N-conformation, occurred wide distributed along the simulation time (2 ns and 37 ns), but the reverse rotation is not observed.
- Flow of water molecules into the FAD pocket, also facing the FAD N_5atom (2 occurrence), while Asn378 stays in the initial O-conformation.

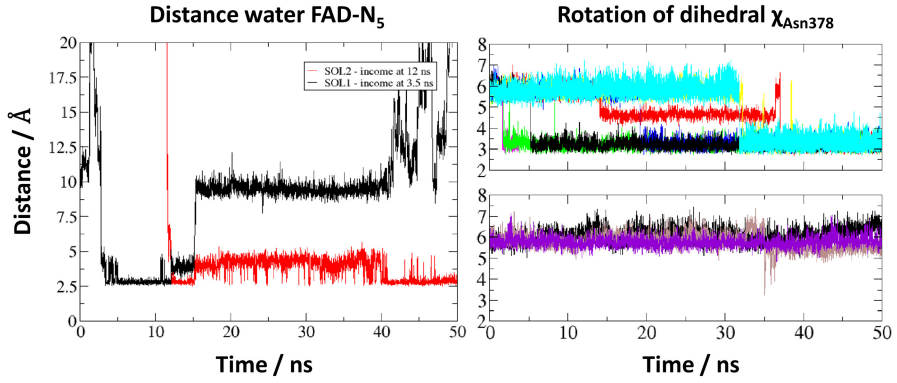

Figure 9.6.: Structral rearangement around $FAD^{\bullet-}$ in 10 MD simulations: the right picture shows the distance between the side chain of Asn378 and FAD N_5in the 10 individual MD simulations showing the reorientation between O- to N-conformation. Top part of the right picture: seven runs in which N-conformation is occupied at the end of the simulation. Bottom part of the right picture: three simulations with stable O-conformation, two with an income of water in the FAD pocket and the one other without water incomes. The distance between the entering water and the FAD N_5 is studied explicitly and shown in the left picture (red and black lines). The movement of water in free MD simulations with $FAD^{\bullet-}$ is indicated by the measured distance between the centre of mas of the waters and the FAD N_5 and a distance of less than 3 Å indicates that the water entered the FAD pocket by passing through a loop.

9.4.1. Rotation of the Asn378 side chain affects protonation of $FAD^{\bullet-}$

Classical MD simulations, having a $FAD^{\bullet-}$ in the pocket, give a first qualitative insight of possible structural reorientations in particular showing the preference of the Asn378 rotation. However, a more detailed understanding of the energetics of this side chain rotation around the C_β-C_γ bond requires further quantifications which can be done by means of biased sampling simulations.

Well-tempered metadynamics simulations are performed to address this conformational reorientation step, while four different states of FAD are considered (FAD_{ox}, $FAD^{\bullet-}$, $FADH^{\bullet}$ and $FADH^{-}$). The free energy profiles of this rotation between the O- to N-conformation presents the energetic preference of the different FAD states, which are illustrated in Figure 9.8. FAD_{ox}, $FADH^{\bullet}$ and $FADH^{-}$ favour the O-conformation by 4-5 kcal·mol^{-1} clearly. On the contrary, the $FAD^{\bullet-}$ state shows a global minimum at the

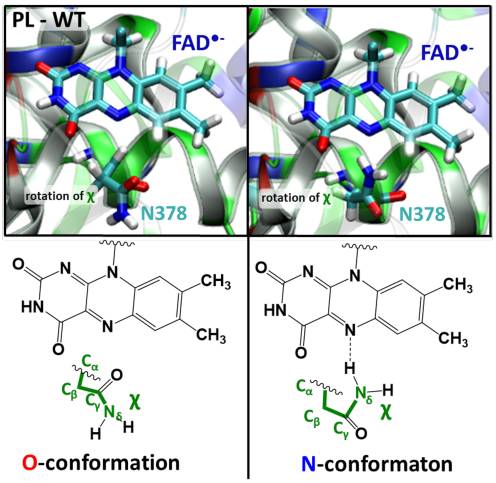

Figure 9.7.: Different orientations of the Asn378 side chain in PL-WT: a view of the O- (left) and N-conformation (right) of Asn378 and its closest protein environment. The dihedral angle χ is constructed out of the N_δ, C_γ, C_β and C_α of Asn378. N-conformation ($\chi = 90\,^\circ$): the amino group of Asn is rotated close to FAD N_5, forming a hydrogen bond. O-conformation ($\chi = 240\,^\circ$): the amido oxygen atom of Asn378 forms a hydrogen bond with a protonated FADH$^\bullet$/FADH$^\bullet$. At the bottom of the picture, the position of FAD and χ are visualised with chemical structures for both conformations.

N-conformation, while the O-conformation is around 3 kcal·mol^{-1} higher and separated by a barrier of 6 kcal·mol^{-1}.

The energy barriers, separating the O- and N-conformation can be correlated to rate constants, which are listed in Table 9.4. The structural motives may explain the energetic preferences:

(I) In case of FAD$^{\bullet-}$ the rotation into the N-conformation may be driven in order to avoid an electrostatic mismatch between the Asn378-oxygen and the negative charge on FAD, and the formation of a hydrogen bond between Asn378 and FAD N_5 can further stabilise the N-conformation (shown in Figure 9.7).

(II) For protonated FAD (FADH$^\bullet$ and FADH$^-$), the N-conformation may be disfavoured due to a steric hindrance between H-N_5 of protonated FAD and the Asn378 side chain.

(III) In the FAD_{ox} state the O-conformation is favoured by about 5 kcal·mol^{-1}, which is surprising since, as in FAD$^{\bullet-}$, a hydrogen bond in the N-conformation could be formed. However, the stabilisation of the O-conformation is due to a hydrogen bonded network (not shown), which can be formed by several water molecules and the Glu363 side chain.

The results are consistent with the experimental and computational studies by Wijaya et al.,[66] which reported a hydrogen bond between the amido oxygen of Asn378 and N_5-H of FADH$^\bullet$ and FADH$^-$ states as well as the absence of the N-conformation in FAD_{ox}.

Table 9.4.: Calculated barrier heights of the rotation of χ in Asn378 for FAD_{ox}, $FAD^{\bullet-}$, $FADH^{\bullet}$, $FADH^{-}$ and the corresponding reaction constants are presented.

	$Barrier_{N\text{-}O\text{-}conformation}$ [kcal·mol^{-1}]	$Barrier_{O\text{-}N\text{-}conformation}$ [kcal·mol^{-1}]	$k_{N\text{-}O}$ [ns^{-1}]	$k_{O\text{-}N}$ [ns^{-1}]
FAD_{ox}	4.67	8.14	2.48	0.01
$FAD^{\bullet-}$	6.12	2.31	0.22	129.76
$FADH^{\bullet}$	1.77	6.14	321.02	0.21
$FADH^{-}$	2.91	6.97	47.43	0.052

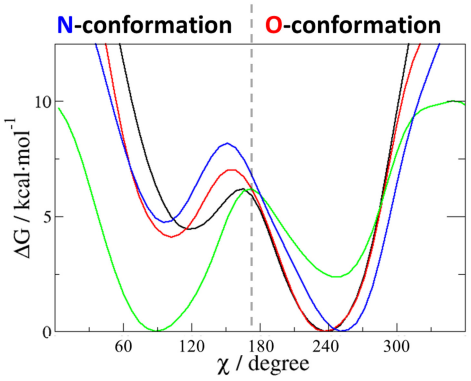

Figure 9.8.: Free energy profile of the rotation of the side chain of Asn378: the different FAD states of PL-WT are shown by different colours (FAD_{ox} (blue), $FAD^{\bullet-}$ (green), $FADH^{\bullet}$ (black) and $FADH^{-}$ (red)). The reaction coordinate is χ the torsion angle of Asn378 (N_{δ}-C_{γ}-C_{β}-C_{α}); 90° corresponds to the N-conformation, while the O-conformation is found at 240°.

9.4.2. The transient formation of a water wire

Immediately after FAD_{ox} photo-reduction to $FAD^{\bullet-}$, the system is in the O-conformation, since this is the favourable state for FAD_{ox}. Then, it is possible to protonate the N_5 of $FAD^{\bullet-}$, however, the energetic study of the Asn378 rotation shows that in presence of the charged $FAD^{\bullet-}$ the rotation into the N-conformation is favoured. The free MD simulations show the occurrence and therefore this rotation may be a dead end for the $FADH^{-}$ formation, since the hydrogen bond hinders protonation. Hence, for the protonation of N_5 the rotation has to be reversed.

As discussed above, the formation of a water wire may connect a proton donor at the protein surface with FAD and thereby enable a path for a PT. In three of the unbiased MD simulations reported above, Asn378 remained in the O-conformation, and in two of them one or two water molecules entered the FAD pocket and engaged in hydrogen bonding with FAD N_5, connecting this cofactor with the protein exterior.

One of these entering events is visualised in the right part of Figure 9.6 and will be analysed in te following. The distances between the water molecules and FAD are presented. At the beginning of the simulation, both water molecules are located in the

bulk solvent. After 2.5 ns, one of them enters the pocket to interact with FAD N_5 and after 10 ns, the other water molecule also enters. Both of them stay between FAD N_5 and the side chain oxygen of Asn378 for about 2.5 ns. Then, the first water is released, and the second water stays in the FAD pocket for the entire remaining simulation time.

The entrance of water molecules into the FAD pocket proceeds through a flexible loop composed of Met367 and Leu376. There are multiple other water molecules located in front of the loop, shown in Figure 9.10, which allow the formation of a water wire connecting $FAD^{\bullet -}$ with the bulk solvent. Importantly, the influx of water molecules into the FAD pocket is vital for the formation of such a wire, which represents a possible pathway for the protonation of FAD.

Furthermore, it could be possible that the entering of the first water changes the character of the pocket and rises the probability for further water to enter or to form the transient wire. To better characterise the kinetics of this event, free energy profiles are computed for the waters entering the FAD pocket. Classical MD simulations in combination with US are used. To reduce the complexity of the possible reorientations in the pocket, the following assumptions are taken: it is focused on just the free energy barrier of the entrance of two water molecules and therefore the rotation of the Asn378 is constrained. Otherwise those two structural reorientations would compete against each other and the convergence of free energy calculations is hard to achieve:
(I.) The first studied reaction focuses just on the entrance of one water, and therefore the coordinate is the distanced between this water molecule and FAD, $d_{FAD-N_5\cdots H_2O}$.
(II.) Afterwards, the interest is in the barrier for a second water to engage with the first water, which would be the first step of a possible wire formation. Thus, the reaction coordinate for the second water is the distance, $d_{[FAD-N_5-H_2O]\cdots H_2O}$, of this water molecule to the first water, which is already fixed in the pocket 3 Å away from FAD. The free energy profiles are shown in Figure 9.9 I. and II.

Taken together, the barrier height for the entrance of the first water (I.) is probably underestimated because the rotation from the N-conformation to the O-conformation cannot occur in the US simulations, which could competes with the entrance of water. Furthermore, the barrier height for the entrance of the second water might be overestimated because the first water is partially constrained in the simulations and cannot interact constructively with the second water to probably improve the entrance. Additionally, the second water is not allowed to penetrate deeper into the pocket and is forced to stay in the energetic unfavoured position in between the loop; the interaction between the two water molecules might thus be underestimated, further, a enclosed protonated H_3O^+ could also change the probability of the entrance.
In summary, this study roughly shows the estimated energetics of a water to enter the pocket and that constructing a wire does not include unphysical high barriers. Further waters are surrounding the loop and could be used to form the wire by extending the two waters easily.

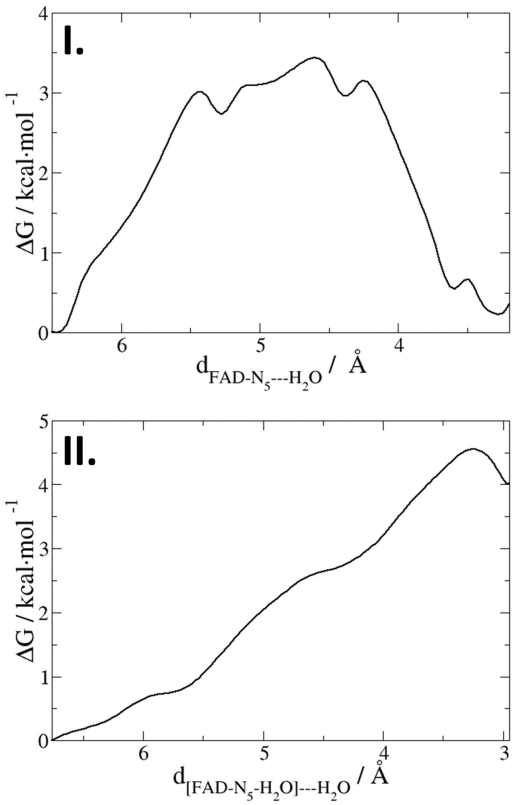

Figure 9.9.: Free energy profiles of the entrance of water molecules into the FAD pocket: I.: one water molecule enters the FAD pocket. II.: an additional water molecule moves into the flexible loop to approach the water molecule being already in the pocket. Reaction coordinates are the distance between FAD and the water molecules (I.) and the distance between the first water and the second water (II.).

9.4.3. Long-range proton transfer along a water wire

The entrance of water molecules allows the formation of a path connecting FAD N_5 with a potential proton donor at the exterior of the FAD pocket. Inspection of the surrounding structure of the pocket identifies Glu106 (with a pKa around 6) as the closest titrable amino acid. Snapshots from the free MD with the two water molecules entering the FAD pocket also show a possible water wire, connecting Glu106 with FAD, as shown in Figure 9.10. Since this water wire is a transient phenomenon, it is necessary to constrain these waters in order to compute the free energy profile for PT along the wire from Glu106 to FAD. Therefore, the water molecules are only allowed to move freely in small spheres around their centre of mass. Biased QM/MM US simulations consider the Glu106 side chain, $FAD^{\bullet -}$ and the five water molecules as possible proton carriers in the QM region. They are performed to estimate the free energy barrier for the PT and the resulting free energy profile is shown in Figure 9.10.

Compared to the direct PT between Asp378 or Asp396 in the previous section, this

PT requires a significantly complexer reaction coordinate. Thus, the construction of the reaction coordinate involves the position of $FAD^{\bullet-}$ N_5, the atoms of the five water molecules forming the wire, and the external proton donor, Glu106. The modified centre of excess charge (mCEC) by König et al.[217] is a useful reaction coordinate for this long-ranged PT:

$$\zeta = \sum_{i=1}^{N_H} r^{H_i} - \sum_{j=1}^{N_X} w^{X_j} r^{X_j} - \sum_{i=1}^{N_N} \sum_{j=1}^{X} f_{sw}(d^{H_i,X_j})(r^{H_i} - r^{X_j}) \tag{9.4}$$

where $r^{(H_i)}$ are the coordinates of all of the hydrogen atoms possibly taking part in the PT process; $r^{(X_J)}$ are the coordinates of the proton acceptors – donor, final acceptor as well as all of the relays, and $w^{(X_j)}$ are the numbers of protons bound to each acceptor in the least protonated state (e.g., w = 2 for a water molecule, H_3O^+). The last term is a correction of coordinate, which runs over all of the hydrogens and proton acceptors and it involves a switching function $f_{sw}(d^{(H_i,X_j))}$ to decide whether each couple of atoms is connected by a bond:

$$f_{sw}(d^{H_i,X_j}) = \frac{1}{(1 + exp[(d^{(H_i,X_j)} - r_{sw})/d_{sw}])} \tag{9.5}$$

Considered are these values of empirical parameters: r_{sw}=1.25 Å and d_{sw}=0.04 Å. The coordinate ξ is a vector quantity that expresses the position of proton being transferred. A scalar reaction coordinate is obtained from ξ as:

$$\zeta = d_{(\xi,D)}/(d_{\xi,D} + d_{\xi,A}) \tag{9.6}$$

here $d_{\xi,D}$ and $d_{\xi,A}$ are the distances from the point ξ to the initial proton donor and to the final proton acceptor, respectively. With such a collective variable, the PT reaction proceeds from ζ=0 for the reactant to ζ=1 for the acceptor.

Analysis of the reaction coordinate, ζ, allows to identify the position of the proton along the reaction, and thereby the connection of the proton movement with transition states and local minima along the reaction. Based on that, the following steps can be identified:

(I) The first barrier of 19 kcal·mol^{-1} is associated with the PT from Glu106 to the water wire. Moving the proton to the first water molecule has an energy penalty of 12.5 kcal·mol^{-1}, while moving the proton further to the second water molecule after crossing the first barrier leads to the fist stable local minimum.

(II) The second step starts with a PT proceeding along the water wire and through the loop, from the third to the fourth water molecule, over another barrier of 7 kcal·mol^{-1}.

(III) The transfer to the fifth water molecule and then to the N_5 atom of FAD is a downhill process resulting in a product state which is 10 kcal·mol^{-1} below the reactant state.

This study shows that the rate limiting step is the deprotonation of the donor which can be assumed to a rate constant of around $0.13\,s^{-1}$, which corresponds to a mean reaction time of 7.7 s. This is in the same order of magnitude than the experimental measurement reaction time of 4 s.[43,221]

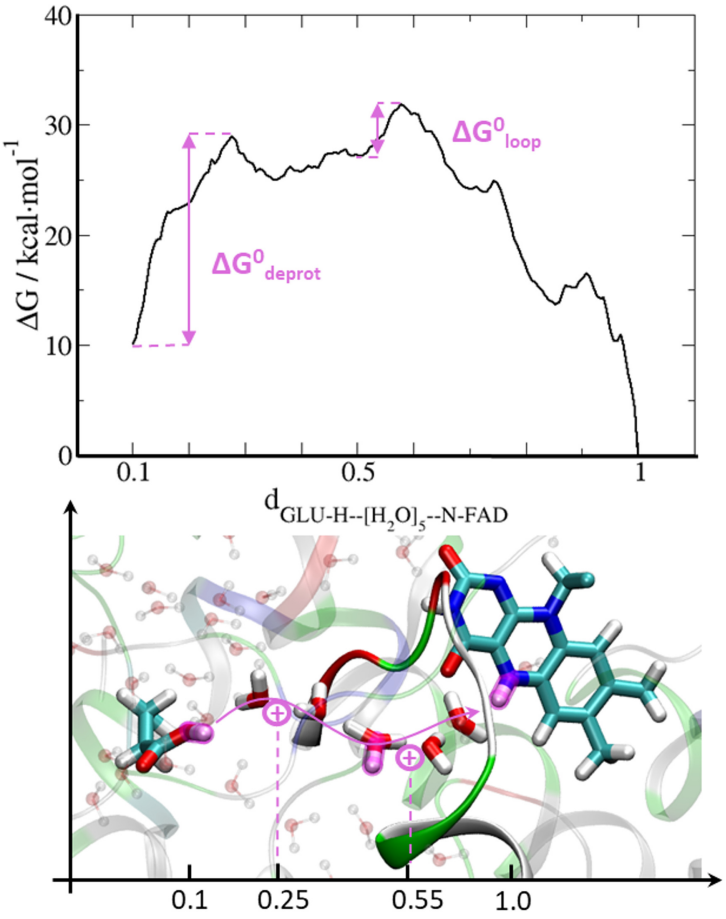

Figure 9.10.: Long range PT in PL-WT: top: free energy profile of the PT along water wire between Glu106 and $FAD^{\bullet -}$. The reaction coordinate is constructed out of the positions of the water molecules, carboxyl group of Glu106 and FAD N_5. Bottom: view of the PT pathway (pink) between Glu106 side chain (at 0.1), *via* water molecules (in the centre), to FAD (at 1.0) which are shown explicitly while the surrounding environment is just transparently presented.

9.4.4. Stability of $FADH^{\bullet}$ vs. $FADH^{-}$ in the different proteins

With respect to the biological functions of CRY or PL, a different protonation state of FAD needs to be stabilised. Thus, CRYs are used as light-sensitive signalling molecules which are able to form a $FAD^{\bullet -}$ – Trp^{+} radical pair, especially plant CRY further form $FADH^{\bullet}$. Nevertheless, they will return in their starting state having FAD_{ox}, after remaining several seconds in the $FAD^{\bullet -}$ state. To fulfil this circle, the availability of a proton

donor and also acceptor needs to be ensured by the nature of the pocket.

On the contrary, Asn allows PL to stabilise a $FADH^{\bullet}$, which is necessary for their biological function. The protonation of $FAD^{\bullet-}$ along the water wire showed a strongly exothermic PT, highlighting the large stabilisation (around 10 $kcal \cdot mol^{-1}$) of $FADH^{\bullet}$ in the PL pocket, while in CRY the difference was just around 2 $kcal \cdot mol^{-1}$.[18] Furthermore, no strong acceptor of the proton would be available in PL to deprotonate $FADH^{\bullet}$, hence, the further photo-induced ET forming $FADH^{-}$ can happen without any competing side reaction.

This addresses the question, why CRYI is not able to form $FADH^{-}$ while it shows the same requested Trp Triad enabling the photo-induced activation of the enzyme. In CRYI, the PT forming $FADH^{\bullet}$ results in a close negative charged Asp396, which can be stabilised by the environment but may still influence the energetics of $FADH^{\bullet}$ in the pocket. Furthermore, its suggested that the negative charge at Asp396 induces conformational changes which may be required for the signal transduction, therefore, the negative charge my be related to the functionality of the protein.[61,195]

To investigate this point, the stabilisation of the $FADH^{-}$ state, induced by the pocket, needs to be calculated. Therefore, the same approach, previously described in chapter 6 and 7 is used to calculated the energy of the HOMO of $FADH^{-}$ in PL-WT, PL-N378D and CRYI; a QM/MM simulation is used to calculate the energy of the selected orbitals, taking the surrounding environment into account. While the electron is fixed at the $FADH^{\bullet}$ and therefore modelling $FADH^{-}$, the MM region is free to adopt to that new charge state. This results in a decrease or increase of the orbital energy, which corresponds to a stabilising o destabilising interaction, respectively.

The simulation of PL-WT shows a decrease of the orbital energy of around 1.48 eV which is 0.36 eV and even 0.69 eV more than obtained from the same simulation approach in PL-N378D and CRYI, respectively. The energy fluctuation of the three systems is also shown in Figure 9.11 and the decrease of orbital energy in PL-WT and PL-N378D is visible, while the orbital energy just slightly fluctuates in CRYI. The close negative charge at the Asp in CRYI and PL-N378D may not allow the same stabilisation than the Asn378 in PL-WT. Due to the similarity of the pocket in PL-WT and PL-N378D it can be concluded that the charge of the Asp378 causes this missing stabilisation. Furthermore, it highlights that the pocket of PL-WT or PL-N378D have a stronger stabilising effect on a negative charge on FAD than the FAD pocket of CRYI.

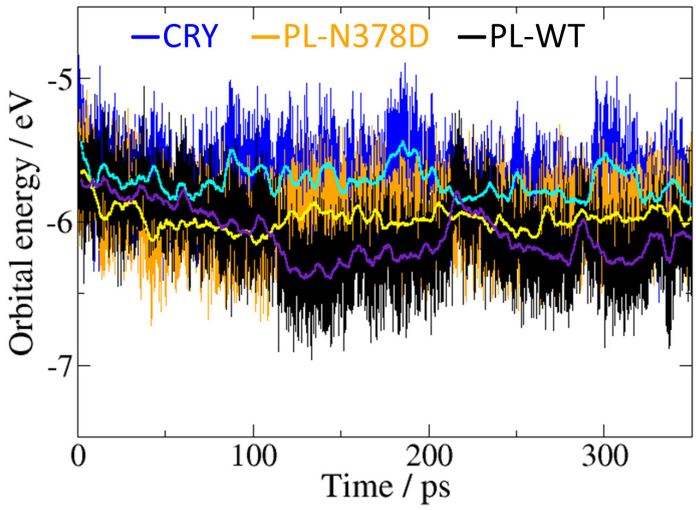

Figure 9.11.: HOMO energy of $FADH^-$ in PL-WT, PL-N378D and CRYI: the energy of the HOMO of $FADH^-$ is calculated in a 350 ps QM/MM MD simulation. The curves show the fluctuation of the orbital energy induced by the environment of the FAD pocket (CRYI blue and averaged value cyan, orange PL-N378D and averaged value yellow and PL-WT black and averaged value purple).

9.5. Conclusion & Discussion

The different function of PL and CRY is directly linked to the occurrence of different FAD oxidation and protonation states, and therefore, one interesting question to be addressed: how are these states are stabilised by the respective protein environments. In this work, classical and QM/MM MD simulations are used to investigate in detail the FAD binding pocket and the factors responsible for the stabilisation and particular PT events. In Figure 9.12, a detailed view of all reported reactions is presented. The experimental measured PT reactions are mentioned and further structural reorientations, addressed in this study, are included to explain the mechanisms on a molecular level.

Experimental studies have identified one crucial amino acid, which is responsible for cofactor protonation in plant CRY (Asp396 in CRYI), while PL have conserves an Asn in this position. This Asn stabilises the biological vital $FADH^\bullet$/$FADH^-$ states in PL, furthermore, it also allows protonation of $FAD^{\bullet -}$. On the contrary, the mutation of Asn into Asp (PL-N378D) prevents the formation of a stabilised $FADH^\bullet$. Additionally, mutating the nearby Glu363, which is an obvious difference between the CRYI and PL-WT, does not change the functionality. But its protonation state has a decisive influence on the PT barrier and deprotonated Glu363 disfavours the PT even more than protonated Glu363. Hence, there are more differences between the CRY and PL active site environments, affecting the architecture of the pocket and thereby influencing the FAD oxidation and protonation states. The initial reorientation of Asp378/396 occurs after a photo-induced

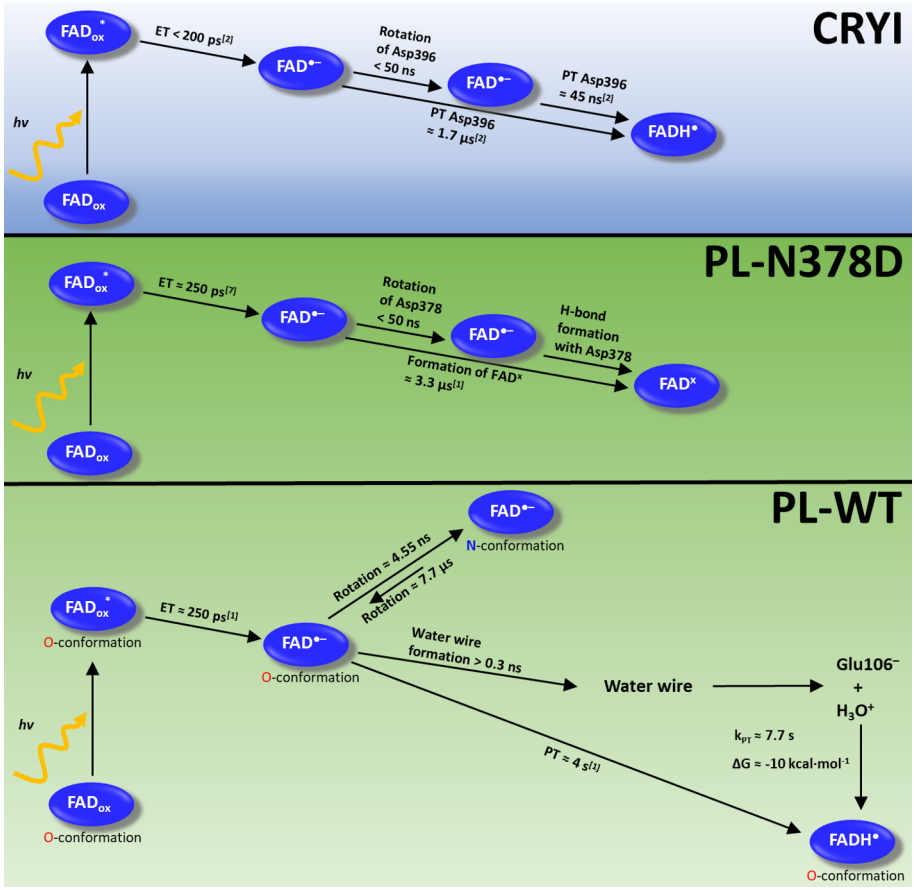

Figure 9.12.: PT and ET reactions in CRYI, PL-N378D and PL-WT: at an experimental level, the ET and PT reactions of PL *E. coli* (green background) were explored by Brettel, indicated with [1],[43] and of CRYI (blue background) by Kottke, indicated with [2].[61,73] The results of biased and unbiased MM and QM/MM simulations are included into the scheme to extend the knowledge about the processes.

ET to FAD (see Figure 9.12) in both proteins, which is essential for a further PT.

Electrostatic interactions with the environment and hydrogen bonding can achieve stabilisation of the FAD oxidation and protonation states. The analysis of both proteins, CRYI and PL-N378D, revealed major differences between the flexibility of the pocket and in the Coulomb interactions, ΔE_{Coul}, between the FAD-Asp ion-pair, which might belong to the different flexibility. The PT energetics are substantially influenced by the rigidity of Arg344 in PL-N378D (Arg362 in CRYI) and the interaction with the water molecules entering the FAD pocket. Furthermore, this rises a different water distribution in CRYI and allows to stabilise the deprotonated Asp396 stronger than in PL-N378D, which results to an exothermic PT in CRYI, while it is endothermic in PL-N378D. In CRYI, a stable $FADH^{\bullet}$ species can be formed, confirmed by experimental and computational studies. On the contrary, the electrostatic interactions in PL-N378D makes this thermodynamically unfavourable, which corresponds to the existence of the FAD^{x} and the absence of the $FADH^{\bullet}$ state. The simulations allow to rise a hypothesis determining the occurrence of the FAD^{x} state: a strong hydrogen-bonding network between Asp378 exists

and therefore, $FADH^{\bullet}$ can be formed transiently but not stabilised, which is indicated by the fast reaction rates, calculated using the biased QM/MM simulations.

Protonation of $FAD^{\bullet-}$ requires a proton donor, which is adjacent to the FAD in plant CRY, while possible donors are located outside of the FAD binding pocket in PL. This can explain the much longer PT times, observed in the experiment. Thus, the simulations suggest a potential proton donor Glu106 outside the FAD binding pocket. Then, the PT to FAD needs the transient formation of a water wire, a step which is endothermic and occurring on a nano second timescale. In addition to that, this wire formation is competing against the rotation of Asn378 (Figure 9.12), which is exothermic in presence of $FAD^{\bullet-}$.

The PT along the wire is exothermic, but has to overcome a large barrier of more than 19 $kcal \cdot mol^{-1}$ corresponding to the timescale in the seconds regime, which is in accordance with the experiment.

Taken together, the long-ranged PT in PL-WT needs to form an energetically unfavourable water wire, while the rotation of an amino acid, Asn378, is an energetically favoured rearrangement. In addition to that, the travelling proton has to overcome a sizeable high barrier, which highlights why this transfer happens on a completely different timescale compared to the fast PT in CRYI.

Therefore, the absence of a titrable amino acid close to FAD requires this complex long-ranged PT which is in general possible in PL-WT. The presence of Asn strongly stabilises protonated $FADH^{\bullet}$ or $FADH^{-}$, highlighted by the exothermic PT, and an efficient PT mechanism is unnecessary because PL just require a stable $FADH^{\bullet}$.

Furthermore, the negative charge on the Asp378/396 would interact with the negative charge on $FADH^{-}$, thereby not allowing a second ET after $FADH^{\bullet}$ has been formed. Hence, the FAD pocket of PL has to avoid a titrable amino acid close to FAD, while the protonation from outside the binding pocked is also still possible to reprotonate FAD after an unexpected deprotonation.

In CRYI on the contrary, the present Asp396 in the FAD pocket allows fast protonation and deprotonation of FAD during the photocycle,[61] therefore, the $FADH^{\bullet}$ state is stabilised while the $FADH^{-}$ state is destabilised by the negative enclosed charge at Asp396.
For DNA repair, in contrast, a $FADH^{-}$ is required, because this it has the proper electronic structure after excitation to allow an ET to the DNA lesion. Therefore, the experimental study mutating the Asp396 in CRYI into Asn interestingly showed the presence of a DNA repair ability, which highlights the important role of Asn enclosed to the $FADH^{\bullet}$.[55]

10. Conclusion

In this work, light-induced charge transfer processes in proteins of the photolyase (PL) and cryptochrome (CRY) family were investigated to solve experimentally raised questions using approximate QM/MM molecular dynamics schemes.

In chapter 6 and 7, an established multi-scale method was used to determine the time-resolved charge transfer (CT) processes in a 6-4-PL, *PhrB*, and a CPD-PL, *PhrA*, respectively. While the photo-activation of both proteins have been examined experimentally, resolving the charge transfer kinetics was not possible, which requires expensive time-resolved spectroscopy. Thus, this work used theoretical simulations to explore the CT kinetics and to resolve further processes which are induced by the CT. This increases the understanding of these proteins and showed the relation between their structure and complex functionality.
PhrB shows an unusually placed tyrosine (Tyr) residue in the CT pathway, where in other members of the protein family a tryptophan (Trp) is conserved. To derive a structure-function relationship between CT kinetics and this Tyr, the wild type and additional mutants, which substitute this Tyr, were analysed. The simulations showed that the solvent distribution around the pathway is different in *PhrB*. Therefore, the energetically unfavoured Tyr introduces an essential barrier to hinder the backward CT and the charge recombination. On the contrary, a mutant, which includes a Trp in the pathway, facilitates the backward transfer. Additionally, the analyses revealed that mutations, containing an alanine instead of the Tyr, did not affect the geometry of the pathway. Thus, experimental requested alternative pathways can be excluded. Calculations of the electronic couplings support that in *PhrB* energetically unfavourable amino acids, which interrupt the CT pathway in other PL, allow CT *via* a superexchange tunnelling mechanism.

PhrA presents beside the conserved pathway an alternative branched CT pathway, which makes the CT kinetics more complex. The transfer rates for both pathways were calculated and different transfer times were observed. The conserved pathway offers stronger electronic couplings to ensure a faster transfer compared with the alternative pathway and with other PL. The kinetics of the alternative pathway showed considerable slower transfer rates, especially for the first step of the transfer. A comparison between both pathways revealed that the alternative pathway offers a better stabilisation of the separated charge states. The environment around the last Trp of the alternative pathway increases the stability of the transiently formed $Trp^{\bullet+}$. Furthermore, the backward transfer along the alternative pathway includes a higher barrier, which delays the re-

combination of the charge.

The multi-scale method successfully calculated the CT properties of members of the photolyase/cryptochrome protein family. However, simulations in other molecular complexes showed a limitation of the applied quantum propagation of the charge.
In chapter 8, a test system, based on an anthracene crystal, was created and used to benchmark the established Ehrenfest propagation method. Additionally, an alternative surface hopping propagation was implemented. The electronic couplings between the molecules were artificially scaled, to influence the CT properties of the system. Both propagators showed coupling-dependent changes in their CT mobility. The unscaled couplings present the expected band-like transfer regime, typical for organic-semiconductors, while the reduced couplings showed a hopping-like transfer. The Ehrenfest propagation reproduced the experimental measured mobility, while the implemented surface hopping underestimated them systematically. However, the charge carrier travelled localised along the system, which was the observed limitation in the Ehrenfest propagation.

Proton transfer (PT) processes or structural rearrangements are often induced by an initial CT. In PL and CRY, the photo-induced CT triggers a structural reorientation of the reaction centre and an additional PT to the FAD cofactor, which is studied in chapter 9. In the *E. coli* PL, the CT forms a negatively charged $FAD^{\bullet -}$ which can be further protonated to $FADH^{\bullet}$, however, the proton donor and mechanism is unclear. Thus, biased MD and QM/MM simulations were used to propose a PT mechanism, involving a transient water wire pathway and glutamate as initial donor. CRY have a donor in their pocket and can protonate FAD million-times faster than the *E. coli* PL. Surprisingly, experimental mutation studies showed that insertion of a proton donor into the FAD pocket hinders the PT. Thus, biased QM/MM simulations addressed this observation and it was determined that the $FAD^{\bullet -}$ state resembles the free energy minimum of the PT in the mutant. The analysis of classical MD simulations presents that the mutation cannot change the functionality, even a double mutant could not achieve it. This revealed that the complex structure-function relationship between two similar proteins cannot be changed by simple mutations. Therefore, the layout of the whole pocket is responsible and optimised to enable the biological functionality.

The applied multi-scale method could be used to clarify experimental issues and further establishes the possibility to simulate CT in proteins as an alternative to the expensive and time-consuming experimental measurements. Additionally, the atomistic insight into the active regions allowed a deeper understanding of the relation between the protein structure and the corresponding biological function.
The evolutionarily developed functionality of this protein family results in different proteins, PL and CRY, which share a similar protein architecture, but have an individually optimised FAD pocket. Tyr and Trp residues institute CT pathways to ensure the re-

quired switching between the FAD oxidation states.
In this work, it was observed that the arrangement of the FAD pocket, the surrounding structures and the solvent accessibility collaborate to allow a specific proton and electron transfer with respect to the individual protein structure.

Part IV.

Appendix

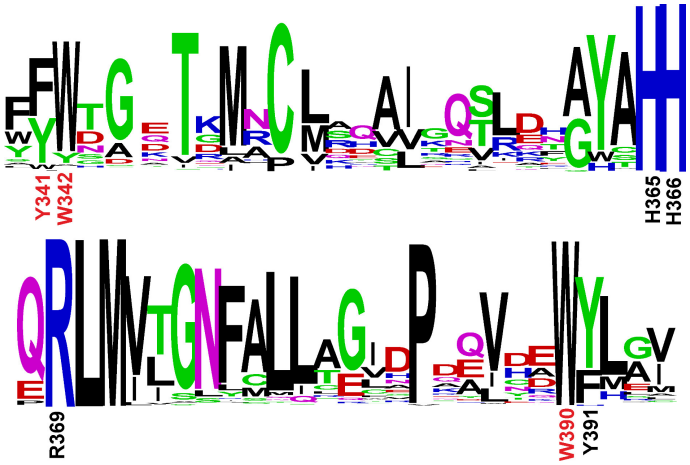

Figure 10.1.: Sequence alignment of the 464 PhrB homologs FeS-BPC proteins using WebLogo web based application: the numbering of the vertical numbers (Y391, W390, W342 etc.) is the amino acid numbering in PhrB and the letters above correspond to the amino acids which are present in the studied proteins at this position. The one-letter-code of the amino acids is used and the size of the letters corresponds to the percentage of the representation in the studied proteins. For example the blue capital H represents that all proteins show a conserved Histidin at a similar position compared to the position His366 in PhrB.

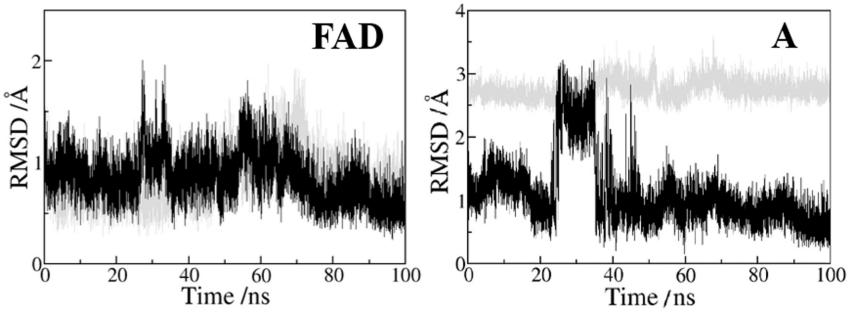

Figure 10.2.: RMSD evolution of FAD in Y391Wd/p: the left shows the RMSD of FAD and the right the RMSD of **A**, along Y391Wp (black) and Y391Wd (grey) simulations (compared to final position of Y391Wp). The highest RMSD values for Y391Wp correspond to the position of **A** closest to **B**, similar to Y391Wd position.

Table 10.1.: Average direct H_{DA} and bridged T_{DA} electronic couplings: FO-DFTB calculations or pathways model was used to obtain the presented couplings with and without the water molecule in the bridge. The pathway model only computes relative values with respect to a reference, i.e. the damping of the charge transfer couplings through covalent bonds, hydrogen bonds and through space.

Electronic Coupling	WT PhrB	Y391F	Y391A
H_{DA} FAD-**B** / meV	0.003	0.005	0.0
T_{DA} FO-DFTB FAD-bridge-**B** / meV	0.062	0.018	-
T_{DA} Pathways backbone (x 10^3) / meV	0.028	0.013	-
T_{DA} Pathways water (x 10^3) / meV	0.835	0.255	0.986

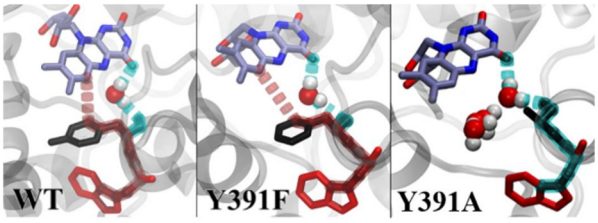

Figure 10.3.: Possible tunnelling pathways: **B** and FAD are connected, which involves **A** and therefore different structures for **A** are studied: aromatic cycle (red) or **A** backbone and a water molecule (cyan) in WT, Y391F and Y391A, respectively.

Table 10.2.: Site energy and electronic couplings of Tyr395 associated to CT pathways including Tyr395 as a charged intermediate (H_{DA}) or as a bridge for tunnelling between FAD and **B** (T_{DA}).

	E(Y395) / eV	H_{DA} FAD-Y395 / meV	H_{DA} Y395-**B** / meV	T_{DA} FAD-**B** / meV
WT	6.19	0.173	17.2	0.020
Y391F	6.26	0.168	10.7	0.007
Y391A	6.11	0.115	11.0	0.015

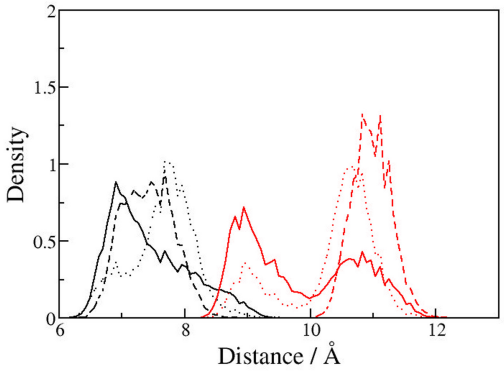

Figure 10.4.: Distribution of the distance between the centres of mass of adenosine and (6-4) thymine (5') (black)- or isoalloxazine and (6-4) thymine (5') (red) for 200 ns of three simulations: $H366^+$ and Mg^{+2} (straight lines), $H366^+$ and Na^+ (dashed lines) and neutral H366 and Mg^{+2} (dotted lines))

Table 10.3.: Electrostatic interactions in kcal/mol between Mg^{2+} cations and 6-4 PL or carboxylate/amide side chain groups of amino acid residues during molecular dynamic simulations. Averaged electrostatic interactions between the Mg^{2+} cations and the 6-4 photoproduct or Asp179, Glu181, Asp254 or the corresponding asparagine after mutation. The interaction between a specific position of Mg^{2+} cation and the 3' phosphate group or the oxygen atom of 3' thymine is averaged along the 1 μs MD simulation using the closest cation for each snapshot. Negative sign means attractive electrostatic interactions while positive sign refers to electrostatic repulsion. The error values correspond to the standard deviation of the interactions during the simulations.

in kcal/mol	6-4-photo-adduct	Asp/ Asn179	Glu181	Asp/ Asn254	Mg_a^{2+}-3'PO_4	Mg_b^{2+}-3'PO_4	Mg_b^{2+}-3'O_2
His366	-18.7 $\pm$ 6.1	-0.1 $\pm$ 7.4	12.3 $\pm$ 7.4	-2.8 $\pm$ 5.3	-	-7.7 $\pm$ 3.6	-5.4 2.8
His366$^+$	-27.9 $\pm$ 5.2	-4.0 +4.7	-3.3 $\pm$ 5.5	-10.0 $\pm$ 5.3	-11.3 $\pm$ 4.1	6.2 $\pm$ 3.7	-6.6 $\pm$ 3.6
D179N$^+$	-21.3 $\pm$ 6.6	1.8 $\pm$ 2.5	-7.4 $\pm$ 7.5	-5.0 $\pm$ 5.7	-9.7 $\pm$ 4.1	4.1 $\pm$ 5.0	-2.8 $\pm$ 3.8
D254N$^+$	-13.3 $\pm$ 5.1	-9.6 $\pm$ 7.9	-9.8 $\pm$ 7.4	-1.0 $\pm$ 2.2	-10.8 $\pm$ 3.7	-	

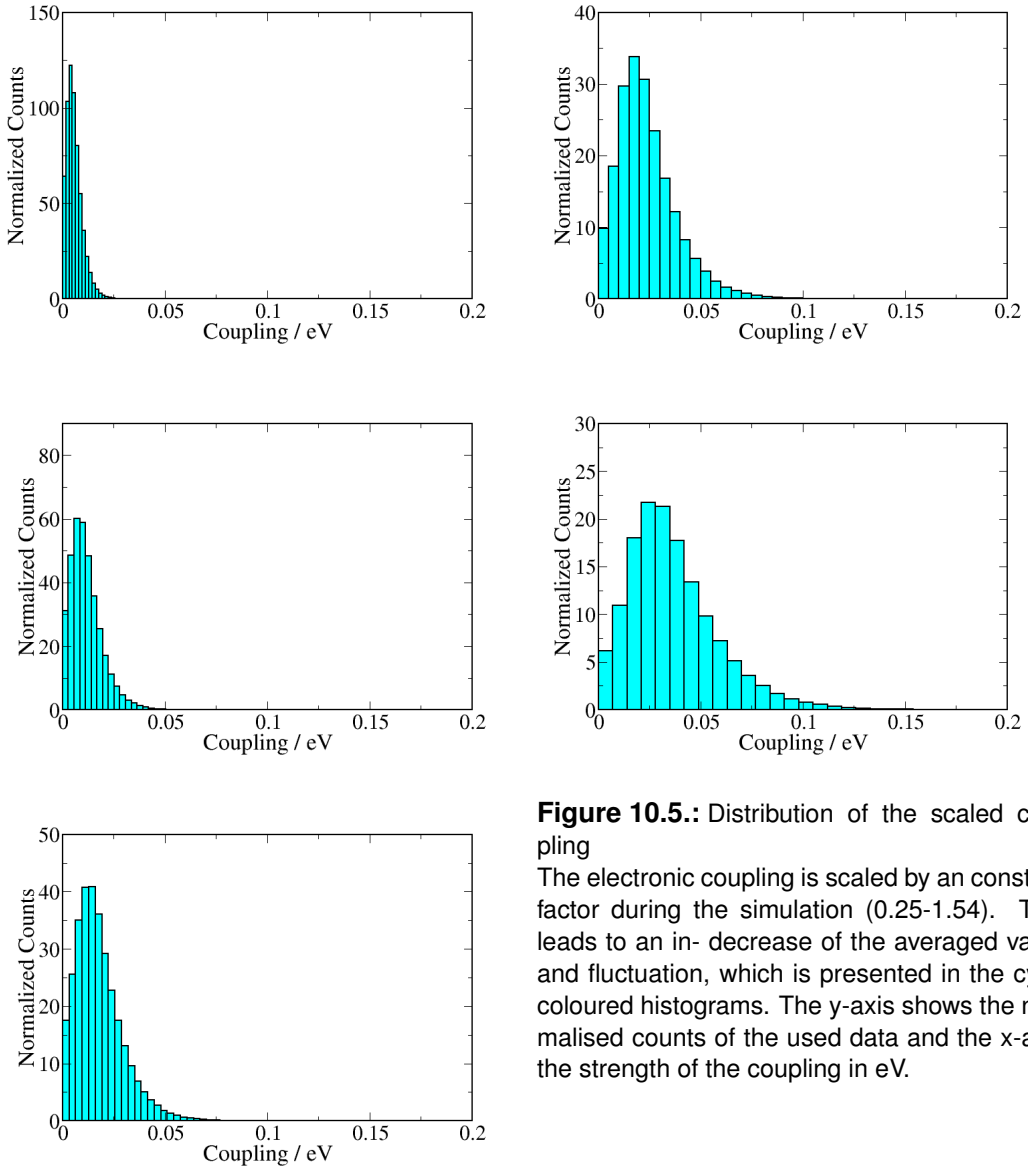

Figure 10.5.: Distribution of the scaled coupling
The electronic coupling is scaled by an constant factor during the simulation (0.25-1.54). This leads to an in- decrease of the averaged value and fluctuation, which is presented in the cyan coloured histograms. The y-axis shows the normalised counts of the used data and the x-axis the strength of the coupling in eV.

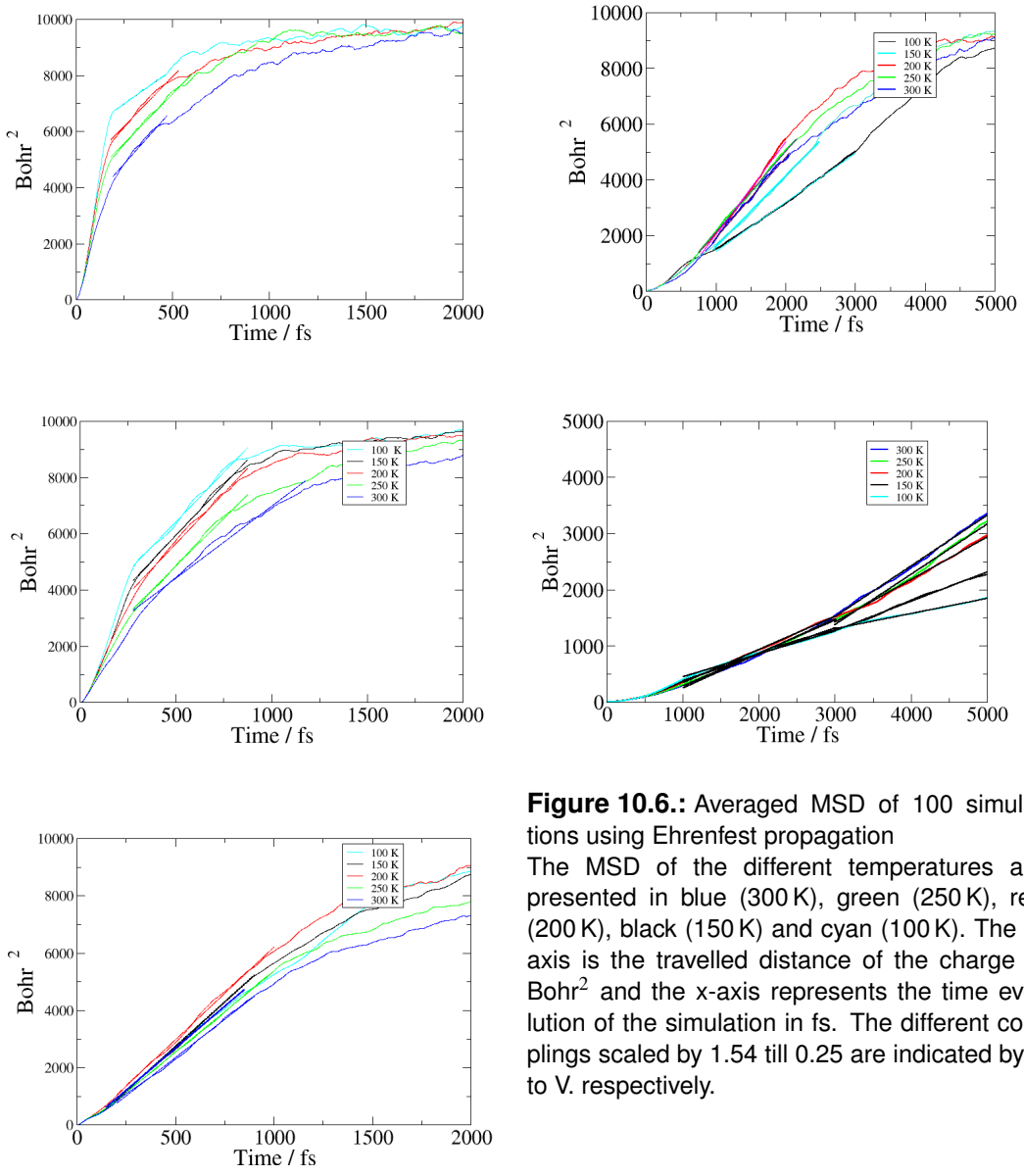

Figure 10.6.: Averaged MSD of 100 simulations using Ehrenfest propagation
The MSD of the different temperatures are presented in blue (300 K), green (250 K), red (200 K), black (150 K) and cyan (100 K). The y-axis is the travelled distance of the charge in Bohr^2 and the x-axis represents the time evolution of the simulation in fs. The different couplings scaled by 1.54 till 0.25 are indicated by I. to V. respectively.

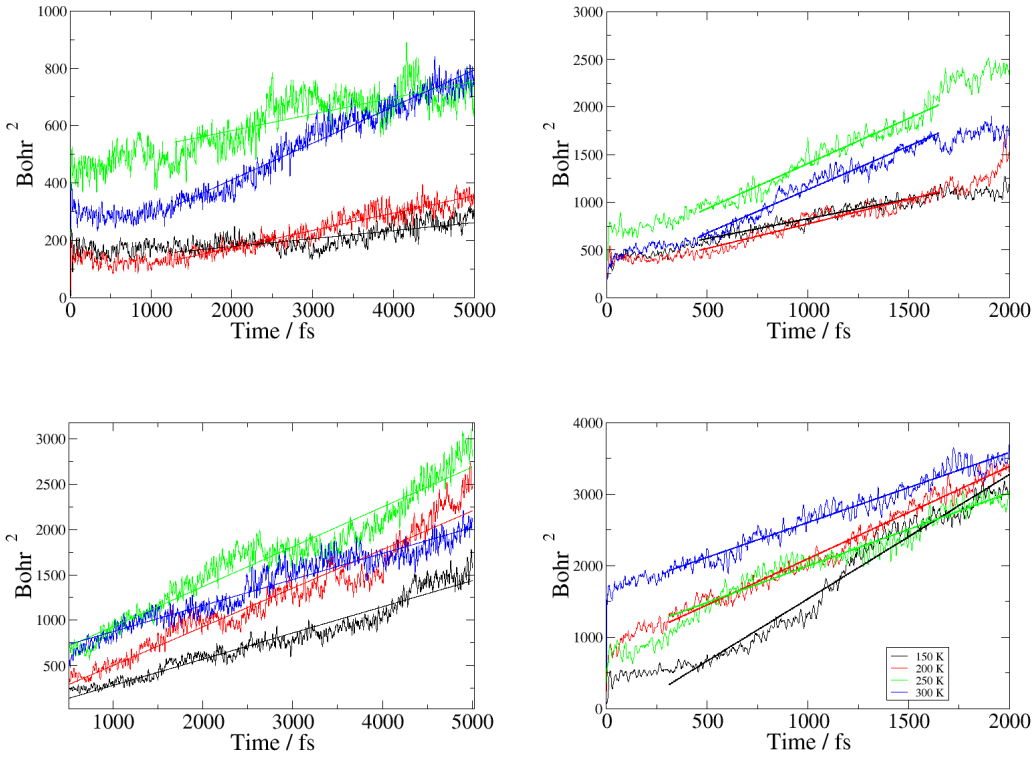

Figure 10.7.: Averaged MSD of 100 simulations using SFH propagation
The MSD of the different temperatures are presented in blue (300 K), green (250 K), red (200 K) and black (150 K). The y-axis is the travelled distance of the charge in Bohr^2 and the x-axis represents the time evolution of the simulation in fs. The different couplings scaled by 1.54 till 0.5 are indicated by I. to IV. respectively.

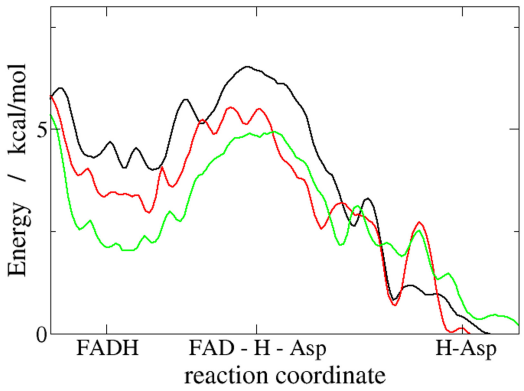

Figure 10.8.: PL-N378D free energy-landscapes for different 363 residues
The correclation between the protonation state of Glu363 are studied and shown. The green curve corresponds to the strongest stabilisation of $FADH^-$ obtained in the presence of protonated Glu363, the red one corresponds to deprotonated Glu363 and the clack one to the mutated Glu363 and presented Met363.

List of Figures

List of Tables

Abbreviations

ATP	Adenosine triphosphate
Ala	Alanine
AS	Amino acids
A. thaliana	Arabidopsis thaliana
AMBER	Assisted Model Building and Energy Refinement
AO	Atomic orbitals
Å	Ångstrom
BO	Born-Oppenheimer-Approximation
CT	Charge transfer
CRY	Cryptochrome
CPD	Cyclobutane-Pyrimidine-Dimer
Cys	Cysteine
DFTB	Density functional tight-binding
DFT	Density functional theory
DNA	Desoxyribo-nuclein-acid
DMRL	6,7-dimethyl-8-ribityllumazine
DIV	Divided frontier charge scheme
EA	Electron affinity
ET	Electron transfer
eV	Electronvolt
E. coli	Escherichia coli
MTHF	Ethenylte-trahydrofolate
fs	Femtosecond
FAD	Flavine adenine dinucleotide
FF	Force fields
FO	Fragment-Orbital
GAFF	General Amber force field
GGA	Generalised gradient approximations
HF	Hartree-Fock
HOMO	Highest occupied molecular orbital
His	Histidine
IP	Ionisation potential
ISO	Isoalloxazin ring system
FeS-BCP	Iron sulphur bacterial Cryptochromes and Photolyases
FeS	Iron sulphur

LZ	Landau–Zener theory
LCAO	linear combination of atomic orbitals
Leu	Leucine
LA	Link-atom
LDA	Local-density-approximation
LC	long-range corrected
LUMO	Lowest unoccupied molecular orbital
Mg	Magnesium
MF	Mean field
μ**s**	Micro seconds
MD	Molecular dynamic
MM	Molecular mechanics
MO	Molecular orbital
MSD	Mean-squared displacement
CT-QM/MM MD	Non-adiabatic molecular dynamics
nm	Nano meter
ns	Nano second
OSC	Organic semiconductor
Phe	Phenylalanine
PCF	Photolyase/Cryptochrome protein family
PL	Photolyase
PhrA	Photolyasen homologues protein A
PhrB	Photolyasen homologues protein B
ps	Pico seconde
PES	Potential energy surface
PMF	Potential of mean force
PriL	Primases
PT	Proton transfer
QM	Quantum mechanic
RESP	Restrained fitting on the electrostatic potential
SE	Schrödinger equation
s	Seconds
SFH	Surface hopping
TS	Transition state
Trp	Tryptophan
Tyr	Tyrosine
US	Umbrella sampling
uv	Ultraviolet
wt	Well-tempered
WHAM	Weighted histogram analysis method
WT	Wild type

Bibliography

[1] Peter Mitchell. Coupling of phosphorylation to electron and hydrogen transfer by a chemi-osmotic type of mechanism. *Nature*, 191(4784):144–148, 1961.

[2] GR Fleming, JL Martin, and J Breton. Rates of primary electron transfer in photosynthetic reaction centres and their mechanistic implications. *Nature*, 333(6169):190, 1988.

[3] Kristina N Ferreira, Tina M Iverson, Karim Maghlaoui, James Barber, and So Iwata. Architecture of the photosynthetic oxygen-evolving center. *Science*, 303(5665):1831–1838, 2004.

[4] Ville RI Kaila. Long-range proton-coupled electron transfer in biological energy conversion: Towards mechanistic understanding of respiratory complex i. *Journal of The Royal Society Interface*, 15(141):20170916, 2018.

[5] Jan Pieter Abrahams, Andrew GW Leslie, René Lutter, and John E Walker. Structure at 2.8 a resolution of f1-atpase from bovine heart mitochondria. *Nature*, 370(6491):621, 1994.

[6] Marten KF Wikstrom. Proton pump coupled to cytochrome c oxidase in mitochondria. *Nature*, 266(5599):271, 1977.

[7] Hannah R Bridges, Andrew JY Jones, Michael N Pollak, and Judy Hirst. Effects of metformin and other biguanides on oxidative phosphorylation in mitochondria. *Biochemical Journal*, 462(3):475–487, 2014.

[8] Charles R Myers and Kenneth H Nealson. Bacterial manganese reduction and growth with manganese oxide as the sole electron acceptor. *Science*, 240(4857):1319–1321, 1988.

[9] Sahand Pirbadian, Sarah E Barchinger, Kar Man Leung, Hye Suk Byun, Yamini Jangir, Rachida A Bouhenni, Samantha B Reed, Margaret F Romine, Daad A Saffarini, Liang Shi, et al. Shewanella oneidensis mr-1 nanowires are outer membrane and periplasmic extensions of the extracellular electron transport components. *Proceedings of the National Academy of Sciences*, 111(35):12883–12888, 2014.

[10] Hans-Werner Fink and Christian Schönenberger. Electrical conduction through dna molecules. *Nature*, 398(6726):407, 1999.

[11] Amie K Boal, Joseph C Genereux, Pamela A Sontz, Jeffrey A Gralnick, Dianne K Newman, and Jacqueline K Barton. Redox signaling between dna repair proteins for efficient lesion detection. *Proceedings of the National Academy of Sciences*, 106(36):15237–15242, 2009.

[12] Spiros S Skourtis. Reviewprobing protein electron transfer mechanisms from the molecular to the cellular length scales. *Peptide Science*, 100(1):82–92, 2013.

[13] Cunlan Guo, Xi Yu, Sivan Refaely-Abramson, Lior Sepunaru, Tatyana Bendikov, Israel Pecht, Leeor Kronik, Ayelet Vilan, Mordechai Sheves, and David Cahen. Tuning electronic transport via hepta-alanine peptides junction by tryptophan doping. *Proceedings of the National Academy of Sciences*, 113(39):10785–10790, 2016.

[14] Jochen Blumberger. Recent Advances in the Theory and Molecular Simulation of Biological Electron Transfer Reactions. *Chemical Reviews*, 115(20):11191–11238, October 2015.

[15] Harald Oberhofer, Karsten Reuter, and Jochen Blumberger. Charge transport in molecular materials: an assessment of computational methods. *Chemical reviews*, 117(15):10319–10357, 2017.

[16] Paul Benjamin Woiczikowski, Thomas Steinbrecher, Tomáš Kubař, and Marcus Elstner. Nonadiabatic QM/MM Simulations of Fast Charge Transfer in Escherichia coli DNA Photolyase. *The Journal of Physical Chemistry B*, 115(32):9846–9863, August 2011.

[17] Gesa Lüdemann, P. Benjamin Woiczikowski, Tomáš Kubař, Marcus Elstner, and Thomas B. Steinbrecher. Charge Transfer in E. coli DNA Photolyase: Understanding Polarization and Stabilization Effects via QM/MM Simulations. *The Journal of Physical Chemistry B*, 117(37):10769–10778, September 2013.

[18] Gesa Lüdemann, Ilia A. Solov'yov, Tomáš Kubař, and Marcus Elstner. Solvent Driving Force Ensures Fast Formation of a Persistent and Well-Separated Radical Pair in Plant Cryptochrome. *Journal of the American Chemical Society*, 137(3):1147–1156, January 2015.

[19] Inês Chaves, Richard Pokorny, Martin Byrdin, Nathalie Hoang, Thorsten Ritz, Klaus Brettel, Lars-Oliver Essen, Gijsbertus T. J. van der Horst, Alfred Batschauer, and Margaret Ahmad. The Cryptochromes: Blue Light Photoreceptors in Plants and Animals. *Annual Review of Plant Biology*, 62(1):335–364, 2011.

[20] Aziz Sancar. Structure and Function of DNA Photolyase and Cryptochrome Blue-Light Photoreceptors. *Chemical Reviews*, 103(6):2203–2238, 2003.

[21] Klaus Brettel and Martin Byrdin. Reaction mechanisms of DNA photolyase. *Current Opinion in Structural Biology*, 20(6):693–701, 2010.

[22] M. S. Jorns, E. T. Baldwin, G. B. Sancar, and A. Sancar. Action mechanism of Escherichia coli DNA photolyase. II. Role of the chromophores in catalysis. *Journal of Biological Chemistry*, 262(1):486–491, May 1987.

[23] G. B. Sancar, M. S. Jorns, G. Payne, D. J. Fluke, C. S. Rupert, and A. Sancar. Action mechanism of Escherichia coli DNA photolyase. III. Photolysis of the enzyme-substrate complex and the absolute action spectrum. *Journal of Biological Chemistry*, 262(1):492–498, May 1987.

[24] Inga Oberpichler, Antonio J. Pierik, Janine Wesslowski, Richard Pokorny, Ran Rosen, Michal Vugman, Fan Zhang, Olivia Neubauer, Eliora Z. Ron, Alfred Batschauer, and Tilman Lamparter. A Photolyase-Like Protein from Agrobacterium tumefaciens with an Iron-Sulfur Cluster. *PLOS ONE*, 6(10):e26775, October 2011.

[25] F. Zhang, P. Scheerer, I. Oberpichler, T. Lamparter, and N. Krauss. Crystal structure of a prokaryotic (6-4) photolyase with an Fe-S cluster and a 6,7-dimethyl-8-ribityllumazine antenna chromophore. *Proceedings of the National Academy of Sciences*, 110(18):7217–7222, 2013.

[26] Daniel Holub, Hongju Ma, Norbert Krauß, Tilman Lamparter, Marcus Elstner, and Natacha Gillet. Functional role of an unusual tyrosine residue in the electron transfer chain of a prokaryotic (6-4) photolyase. *Chem. Sci.*, 9:1259–1272, 2018.

[27] Hongju Ma, Daniel Holub, Natacha Gillet, Gero Kaeser, Katharina Thoulass, Marcus Elstner, Norbert Krauß, and Tilman Lamparter. Two aspartate residues close to the lesion binding site of agrobacterium (6-4) photolyase are required for mg2+ stimulation of dna repair. *The FEBS Journal*, 0(ja).

[28] Satoru Kanai, Reiko Kikuno, Hiroyuki Toh, Haruko Ryo, and Takeshi Todo. Molecular evolution of the photolyase–blue-light photoreceptor family. *Journal of Molecular Evolution*, 45(5):535–548, Nov 1997.

[29] Takeshi Todo. Functional diversity of the dna photolyase/blue light receptor family. *Mutation Research/DNA Repair*, 434(2):89 – 97, 1999.

[30] Jiang Li, Takeshi Uchida, Takeshi Todo, and Teizo Kitagawa. Similarities and differences between cyclobutane pyrimidine dimer photolyase and (6-4) photolyase as revealed by resonance Raman spectroscopy: Electron transfer from the FAD cofactor to ultraviolet-damaged DNA. *The Journal of Biological Chemistry*, 281(35):25551–25559, September 2006.

[31] Ludovic Sauguet, Sebastian Klinge, Rajika L Perera, Joseph D Maman, and Luca Pellegrini. Shared active site architecture between the large subunit of eukaryotic primase and dna photolyase. *PLoS One*, 5(4):e10083, 2010.

[32] Malcolm F White and Mark S Dillingham. Iron–sulphur clusters in nucleic acid processing enzymes. *Current opinion in structural biology*, 22(1):94–100, 2012.

[33] Sebastian Klinge, Judy Hirst, Joseph D Maman, Torsten Krude, and Luca Pellegrini. An iron-sulfur domain of the eukaryotic primase is essential for rna primer synthesis. *Nature Structural and Molecular Biology*, 14(9):875, 2007.

[34] Dominik Graf, Janine Wesslowski, Hongju Ma, Patrick Scheerer, Norbert Krauß, Inga Oberpichler, Fan Zhang, and Tilman Lamparter. Key Amino Acids in the Bacterial (6-4) Photolyase PhrB from Agrobacterium fabrum. *PLOS ONE*, 10(10):e0140955, October 2015.

[35] Hong-Quan Yang, Ying-Jie Wu, Ru-Hang Tang, Dongmei Liu, Yan Liu, and Anthony R Cashmore. The c termini of arabidopsis cryptochromes mediate a constitutive light response. *Cell*, 103(5):815–827, 2000.

[36] Hongju Ma, Fan Zhang, Elisabeth Ignatz, Martin Suehnel, Peng Xue, Patrick Scheerer, Lars Oliver Essen, Norbert Krauß, and Tilman Lamparter. Divalent Cations Increase DNA Repair Activities of Bacterial (6-4) Photolyases. *Photochemistry and Photobiology*, 93(1):323–330, January 2017.

[37] C. Friedberg Errol, C. Walker Graham, Siede Wolfram, D. Wood Richard, A. Schultz Roger, and Ellenberger Tom. *DNA Repair and Mutagenesis, Second Edition*. American Society of Microbiology, January 2006.

[38] Estimates of ozone depletion and skin cancer incidence to examine the Vienna Convention achievements | Nature.

[39] Dongping Zhong. Electron Transfer Mechanisms of DNA Repair by Photolyase. *Annual Review of Physical Chemistry*, 66(1):691–715, 2015.

[40] Stefan Weber. Light-driven enzymatic catalysis of DNA repair: a review of recent biophysical studies on photolyase. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 1707(1):1 – 23, 2005.

[41] Evelyne Sage. Distribution and repair of photolesions in dna: genetic consequences and the role of sequence context. *Photochemistry and photobiology*, 57(1):163–174, 1993.

[42] Daichi Yamada, Hisham M. Dokainish, Tatsuya Iwata, Junpei Yamamoto, Tomoko Ishikawa, Takeshi Todo, Shigenori Iwai, Elizabeth D. Getzoff, Akio Kitao, and Hideki Kandori. Functional conversion of cpd and (6–4) photolyases by mutation. *Biochemistry*, 55(30):4173–4183, 2016.

[43] Pavel Müller, Klaus Brettel, Laszlo Grama, Miklos Nyitrai, and Andras Lukacs. Photochemistry of Wild-Type and N378d Mutant E. coli DNA Photolyase with Oxidized FAD Cofactor Studied by Transient Absorption Spectroscopy. *ChemPhysChem*, 17(9):1329–1340, May 2016.

[44] Patrick Emery, W Venus So, Maki Kaneko, Jeffrey C Hall, and Michael Rosbash. Cry, a drosophila clock and light-regulated cryptochrome, is a major contributor to circadian rhythm resetting and photosensitivity. *Cell*, 95(5):669–679, 1998.

[45] Wolfgang Wiltschko and Roswitha Wiltschko. Migratory orientation: Magnetic compass orientation of garden warblers (sylvia borin) after a simulated crossing of the magnetic equator. *Ethology*, 91(1):70–74, 1992.

[46] Henrik Mouritsen, Ulrike Janssen-Bienhold, Miriam Liedvogel, Gesa Feenders, Julia Stalleicken, Petra Dirks, and Reto Weiler. Cryptochromes and neuronal-activity markers colocalize in the retina of migratory birds during magnetic orientation. *Proceedings of the National Academy of Sciences*, 101(39):14294–14299, 2004.

[47] Anja Günther, Angelika Einwich, Emil Sjulstok, Regina Feederle, Petra Bolte, Karl-Wilhelm Koch, Ilia A. Solov'yov, and Henrik Mouritsen. Double-cone localization and seasonal expression pattern suggest a role in magnetoreception for european robin cryptochrome 4. *Current Biology*, 28(2):211 – 223.e4, 2018.

[48] Meng Chen, Joanne Chory, and Christian Fankhauser. Light signal transduction in higher plants. *Annual Review of Genetics*, 38:87–117, 2004.

[49] Chentao Lin. Blue light receptors and signal transduction. *The Plant Cell*, 14(suppl 1):S207–S225, 2002.

[50] Khushbeer Malhotra, Sang-Tae Kim, Alfred Batschauer, Lale Dawut, and Aziz Sancar. Putative blue-light photoreceptors from arabidopsis thaliana and sinapis alba with a high degree of sequence homology to dna photolyase contain the two photolyase cofactors but lack dna repair activity. *Biochemistry*, 34(20):6892–6899, 1995.

[51] Jean-Pierre Bouly, Baldissera Giovani, Armin Djamei, Markus Mueller, Anke Zeugner, Elizabeth A Dudkin, Alfred Batschauer, and Margaret Ahmad. Novel atp-binding and autophosphorylation activity associated with arabidopsis and human cryptochrome-1. *European journal of biochemistry*, 270(14):2921—2928, July 2003.

[52] Fabien Cailliez, Pavel Müller, Michaël Gallois, and Aurélien de la Lande. Atp binding and aspartate protonation enhance photoinduced electron transfer in plant cryptochrome. *Journal of the American Chemical Society*, 136(37):12974–12986, 2014.

[53] Christian Thöing, Sabine Oldemeyer, and Tilman Kottke. Microsecond Deprotonation of Aspartic Acid and Response of the α/β Subdomain Precede C-Terminal Signaling in the Blue Light Sensor Plant Cryptochrome. *Journal of the American Chemical Society*, 137(18):5990–5999, 2015.

[54] Jing Wang, Xianli Du, Weisong Pan, Xiaojie Wang, and Wenjian Wu. Photoactivation of the cryptochrome/photolyase superfamily. *Journal of Photochemistry and Photobiology C: Photochemistry Reviews*, 22:84–102, March 2015.

[55] Sarah Burney, Ringo Wenzel, Tilman Kottke, Thomas Roussel, Nathalie Hoang, Jean-Pierre Bouly, Robert Bittl, Joachim Heberle, and Margaret Ahmad. Single Amino Acid Substitution Reveals Latent Photolyase Activity in Arabidopsis cry1. *Angewandte Chemie International Edition*, 51(37):9356–9360, 2012.

[56] Erin N. Worthington, İ. Halil Kavakli, Gloria Berrocal-Tito, Bruce E. Bondo, and Aziz Sancar. Purification and Characterization of Three Members of the Photolyase/Cryptochrome Family Blue-light Photoreceptors from *Vibrio cholerae*. *Journal of Biological Chemistry*, 278(40):39143–39154, October 2003.

[57] Haisun Zhu, Quan Yuan, Adriana D. Briscoe, Oren Froy, Amy Casselman, and Steven M. Reppert. The two CRYs of the butterfly. *Current biology: CB*, 15(23):R953–954, December 2005.

[58] Tatsuya Iwata, Yu Zhang, Kenichi Hitomi, Elizabeth D. Getzoff, and Hideki Kandori. Key Dynamics of Conserved Asparagine in a Cryptochrome/Photolyase Family Protein by Fourier Transform Infrared Spectroscopy. *Biochemistry*, 49(41):8882–8891, October 2010.

[59] H. W. Park, S. T. Kim, A. Sancar, and J. Deisenhofer. Crystal structure of DNA photolyase from Escherichia coli. *Science (New York, N.Y.)*, 268(5219):1866–1872, June 1995.

[60] Alexandra Mees, Tobias Klar, Petra Gnau, Ulrich Hennecke, Andre P. M. Eker, Thomas Carell, and Lars-Oliver Essen. Crystal Structure of a Photolyase Bound to a CPD-Like DNA Lesion After in Situ Repair. *Science*, 306(5702):1789–1793, December 2004.

[61] Anika Hense, Elena Herman, Sabine Oldemeyer, and Tilman Kottke. Proton Transfer to Flavin Stabilizes the Signaling State of the Blue Light Receptor Plant Cryptochrome. *Journal of Biological Chemistry*, 290(3):1743–1751, January 2015.

[62] Alex Berndt, Tilman Kottke, Helena Breitkreuz, Radovan Dvorsky, Sven Hennig, Michael Alexander, and Eva Wolf. A novel photoreaction mechanism for the circadian blue light photoreceptor drosophila cryptochrome. *Journal of Biological Chemistry*, 282(17):13011–13021, 2007.

[63] Nuri Öztürk, Sang-Hun Song, Christopher P. Selby, and Aziz Sancar. Animal Type 1 Cryptochromes ANALYSIS OF THE REDOX STATE OF THE FLAVIN COFACTOR BY SITE-DIRECTED MUTAGENESIS. *Journal of Biological Chemistry*, 283(6):3256–3263, August 2008.

[64] Ilia A. Solov'yov, Tatiana Domratcheva, and Klaus Schulten. Separation of photo-induced radical pair in cryptochrome to a functionally critical distance. *Scientific Reports*, 4:3845, 2014.

[65] Lei Xu, Wanmeng Mu, Yanwei Ding, Zhaofeng Luo, Qingkai Han, Fuyong Bi, Yuzhen Wang, and Qinhua Song. Active Site of Escherichia coli DNA Photolyase: Asn378 Is Crucial both for Stabilizing the Neutral Flavin Radical Cofactor and for DNA Repair. *Biochemistry*, 47(33):8736–8743, August 2008.

[66] I. Mahaputra M. Wijaya, Tatiana Domratcheva, Tatsuya Iwata, Elizabeth D. Getzoff, and Hideki Kandori. Single Hydrogen Bond Donation from Flavin N5 to Proximal Asparagine Ensures FAD Reduction in DNA Photolyase. *J. Am. Chem. Soc.*, 138(13):4368–4376, April 2016.

[67] Carol L Thompson and Aziz Sancar. Photolyase/cryptochrome blue-light photoreceptors use photon energy to repair DNA and reset the circadian clock. *Oncogene*, 21:9043, December 2002.

[68] Jie Gao, Xu Wang, Meng Zhang, Mingdi Bian, Weixian Deng, Zecheng Zuo, Zhenming Yang, Dongping Zhong, and Chentao Lin. Trp triad-dependent rapid photoreduction is not required for the function of arabidopsis cry1. *Proceedings of the National Academy of Sciences*, 112(29):9135–9140, 2015.

[69] Patrick Scheerer, Fan Zhang, Jacqueline Kalms, David von Stetten, Norbert Krauß, Inga Oberpichler, and Tilman Lamparter. The Class III Cyclobutane Pyrimidine Dimer Photolyase Structure Reveals a New Antenna Chromophore Binding Site and Alternative Photoreduction Pathways. *Journal of Biological Chemistry*, 290(18):11504–11514, January 2015.

[70] Christopher Engelhard, Xuecong Wang, David Robles, Julia Moldt, Lars-Oliver Essen, Alfred Batschauer, Robert Bittl, and Margaret Ahmad. Cellular metabolites enhance the light sensitivity of arabidopsis cryptochrome through alternate electron transfer pathways. *The Plant Cell*, 26(11):4519–4531, 2014.

[71] Zheyun Liu, Chuang Tan, Xunmin Guo, Jiang Li, Lijuan Wang, Aziz Sancar, and Dongping Zhong. Determining complete electron flow in the cofactor photoreduction of oxidized photolyase. *Proceedings of the National Academy of Sciences*, 110(32):12966–12971, 2013.

[72] Zheyun Liu, Chuang Tan, Xunmin Guo, Jiang Li, Lijuan Wang, Aziz Sancar, and Dongping Zhong. Determining complete electron flow in the cofactor photoreduction of oxidized photolyase. *Proceedings of the National Academy of Sciences*, 110(32):12966–12971, 2013.

[73] Dominik Immeln, Alexander Weigel, Tilman Kottke, and J. Luis Pérez Lustres. Primary events in the blue light sensor plant cryptochrome: Intraprotein electron and proton transfer revealed by femtosecond spectroscopy. *Journal of the American Chemical Society*, 134(30):12536–12546, 2012.

[74] Dominik Immeln, Alexander Weigel, Tilman Kottke, and J. Luis Pérez Lustres. Primary events in the blue light sensor plant cryptochrome: Intraprotein electron and proton transfer revealed by femtosecond spectroscopy. *Journal of the American Chemical Society*, 134(30):12536–12546, 2012.

[75] Slobodan V. Jovanovic, Anthony Harriman, and Michael G. Simic. Electron-transfer reactions of tryptophan and tyrosine derivatives. *The Journal of Physical Chemistry*, 90(9):1935–1939, April 1986.

[76] Stephan Kiontke, Yann Geisselbrecht, Richard Pokorny, Thomas Carell, Alfred Batschauer, and Lars-Oliver Essen. Crystal structures of an archaeal class II DNA photolyase and its complex with UV-damaged duplex DNA. *The EMBO Journal*, 30(21):4437–4449, November 2011.

[77] Till Biskup, Bernd Paulus, Asako Okafuji, Kenichi Hitomi, Elizabeth D. Getzoff, Stefan Weber, and Erik Schleicher. Variable Electron Transfer Pathways in an Amphibian Cryptochrome TRYPTOPHAN VERSUS TYROSINE-BASED RADICAL PAIRS. *Journal of Biological Chemistry*, 288(13):9249–9260, March 2013.

[78] Christopher Engelhard, Xuecong Wang, David Robles, Julia Moldt, Lars-Oliver Essen, Alfred Batschauer, Robert Bittl, and Margaret Ahmad. Cellular Metabolites Enhance the Light Sensitivity of Arabidopsis Cryptochrome through Alternate Electron Transfer Pathways. *The Plant Cell*, 26(11):4519–4531, November 2014.

[79] Andras Lukacs, André PM Eker, Martin Byrdin, Sandrine Villette, Jie Pan, Klaus Brettel, and Marten H Vos. Role of the middle residue in the triple tryptophan electron transfer chain of dna photolyase: ultrafast spectroscopy of a trp phe mutant. *The Journal of Physical Chemistry B*, 110(32):15654–15658, 2006.

[80] Corinne Aubert, Paul Mathis, André P. M. Eker, and Klaus Brettel. Intraprotein electron transfer between tyrosine and tryptophan in DNA photolyase from Anacystis nidulans. *Proceedings of the National Academy of Sciences*, 96(10):5423–5427, November 1999.

[81] Bridgette A. Barry. Reaction dynamics and proton coupled electron transfer: Studies of tyrosine-based charge transfer in natural and biomimetic systems. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 1847(1):46–54, January 2015.

[82] P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Phys. Rev.*, 136:B864–B871, Nov 1964.

[83] Frank Jensen. *Introduction to Computational Chemistry*. John Wiley & Sons, Inc., USA, 2006.

[84] L. H. Thomas. The calculation of atomic fields. *Mathematical Proceedings of the Cambridge Philosophical Society*, 23(5):542–548, 1927.

[85] E. Fermi. Statistical method to determine some properties of atoms. *Phys. Rev.*, 6:602–607, 1927.

[86] *A Chemist's Guide to Density Functional Theory, 2nd Edition*. Wiley-VCH, 2 edition, July 2001.

[87] W. Kohn and L. J. Sham. Self-consistent equations including exchange and correlation effects. *Phys. Rev.*, 140:A1133–A1138, Nov 1965.

[88] Chengteh Lee, Weitao Yang, and Robert G. Parr. Development of the colle-salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B*, 37:785–789, Jan 1988.

[89] Chabalowski C. F. Frisch M. J. Stephens P. J., Devlin F. J. Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields. *J. Phys. Chem.*, 58(8):11623–11627, 1994.

[90] Axel D Becke. Density-functional exchange-energy approximation with correct asymptotic behavior. *Physical review A*, 38(6):3098, 1988.

[91] Michael G. Medvedev, Ivan S. Bushmarinov, Jianwei Sun, John P. Perdew, and Konstantin A. Lyssenko. Density functional theory is straying from the path toward the exact functional. *Science*, 355(6320):49–52, 2017.

[92] Yingkai Zhang and Weitao Yang. Comment on "Generalized Gradient Approximation Made Simple". *Physical Review Letters*, 80(4):890–890, January 1998.

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

[94] Axel D. Becke. Density-functional thermochemistry. III. The role of exact exchange. *The Journal of Chemical Physics*, 98(7):5648–5652, 1993.

[95] S. H. Vosko, L. Wilk, and M. Nusair. Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. *Canadian Journal of Physics*, 58(8):1200–1211, 1980.

[96] Jeng-Da Chai and Martin Head-Gordon. Long-range corrected hybrid density functionals with damped atom–atom dispersion corrections. *Phys. Chem. Chem. Phys.*, 10:6615–6620, 2008.

[97] J. F. Janak. Janax therorem rework. *Phys. Rev. B*, 18:7165–7168, Dec 1978.

[98] Peter Politzer and Fakher Abu-Awwad. A comparative analysis of hartree-fock and kohn-sham orbital energies. *Theoretical Chemistry Accounts*, 99(2):83–87, Apr 1998.

[99] Michael Gaus, Qiang Cui, and Marcus Elstner. DFTB3: Extension of the Self-Consistent-Charge Density-Functional Tight-Binding Method (SCC-DFTB). *Journal of Chemical Theory and Computation*, 7(4):931–948, 2011.

[100] Lyuben Zhechkov, Thomas Heine, Serguei Patchkovskii, Gotthard Seifert, and Helio A. Duarte. An efficient a posteriori treatment for dispersion interaction in density-functional-based tight binding. *Journal of Chemical Theory and Computation*, 1(5):841–847, 2005.

[101] G. Seifert, D. Porezag, and Th. Frauenheim. Calculations of molecules, clusters, and solids with a simplified lcao-dft-lda scheme. *International Journal of Quantum Chemistry*, 58(2):185–192, 1996.

[102] D. Porezag, Th. Frauenheim, Th. Köhler, G. Seifert, and R. Kaschner. Construction of tight-binding-like potentials on the basis of density-functional theory: Application to carbon. *Physical Review B*, 51(19):12947–12957, 1995.

[103] H. Eschrig. *Optimized LCAO Method and the Electronic Structure of Extended Systems*. 1989.

[104] M. Elstner. SCC-DFTB: What Is the Proper Degree of Self-Consistency? *The Journal of Physical Chemistry A*, 111(26):5614–5621, July 2007.

[105] Marcus Elstner, Dirk Porezag, G. Jungnickel, J. Elsner, M. Haugk, Th Frauenheim, Sandor Suhai, and Gotthard Seifert. Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties. *Physical Review B*, 58(11):7260, 1998.

[106] Michael Gaus, Haiyun Jin, Darren Demapan, Anders S. Christensen, Puja Goyal, Marcus Elstner, and Qiang Cui. DFTB3 Parametrization for Copper: The Importance of Orbital Angular Momentum Dependence of Hubbard Parameters. *Journal of Chemical Theory and Computation*, 11(9):4205–4219, September 2015.

[107] Michael Gaus, Albrecht Goez, and Marcus Elstner. Parametrization and benchmark of dftb3 for organic molecules. *Journal of Chemical Theory and Computation*, 9(1):338–354, 2013.

[108] J. v. Neumann. Physical applications of the ergodic hypothesis. *Proceedings of the National Academy of Sciences*, 18(3):263–266, 1932.

[109] J. v. Neumann. Proof of the quasi-ergodic hypothesis. *Proceedings of the National Academy of Sciences*, 18(1):70–82, 1932.

[110] George D. Birkhoff. Proof of the ergodic theorem. *Proceedings of the National Academy of Sciences*, 17(12):656–660, 1931.

[111] RW Hockney, SP Goel, and JW Eastwood. Quiet high-resolution computer models of a plasma. *Journal of Computational Physics*, 14(2):148–158, 1974.

[112] D. J. Evans and B. L. Holian. The Nose–Hoover thermostat. *The Journal of Chemical Physics*, 83(8):4069, 1985.

[113] M. Parrinello and A. Rahman. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics*, 52(12):7182–7190, December 1981.

[114] PH König, M Hoffmann, Th Frauenheim, and Q Cui. A critical evaluation of different qm/mm frontier treatments with scc-dftb as the qm method. *The Journal of Physical Chemistry B*, 109(18):9082–9095, 2005.

[115] Alessandro Laio and Michele Parrinello. Escaping free-energy minima. *Proceedings of the National Academy of Sciences*, 99(20):12562–12566, 2002.

[116] Alessandro Laio, Antonio Rodriguez-Fortea, Francesco Luigi Gervasio, Matteo Ceccarelli, and Michele Parrinello. Assessing the accuracy of metadynamics. *The Journal of Physical Chemistry B*, 109(14):6714–6721, 2005.

[117] Alessandro Barducci, Giovanni Bussi, and Michele Parrinello. Well-tempered metadynamics: A smoothly converging and tunable free-energy method. *Phys. Rev. Lett.*, 100:020603, Jan 2008.

[118] Alexander Heck, P. Benjamin Woiczikowski, Tomáš Kubař, Kai Welke, Thomas Niehaus, Bernd Giese, Spiros Skourtis, Marcus Elstner, and Thomas B. Steinbrecher. Fragment Orbital Based Description of Charge Transfer in Peptides Including Backbone Orbitals. *The Journal of Physical Chemistry B*, 118(16):4261–4272, April 2014.

[119] Natacha Gillet, Laura Berstis, Xiaojing Wu, Fruzsina Gajdos, Alexander Heck, Aurélien de la Lande, Jochen Blumberger, and Marcus Elstner. Electronic Coupling Calculations for Bridge-Mediated Charge Transfer Using Constrained Density Functional Theory (CDFT) and Effective Hamiltonian Approaches at the Density Functional Theory (DFT) and Fragment-Orbital Density Functional Tight Binding (FODFTB) Level. *Journal of Chemical Theory and Computation*, 12(10):4793–4805, October 2016.

[120] LD Zusman. Outer-sphere electron transfer in polar solvents. *Chemical Physics*, 49(2):295–304, 1980.

[121] Arieh Warshel. Dynamics of reactions in polar solvents. semiclassical trajectory studies of electron-transfer and proton-transfer reactions. *The Journal of Physical Chemistry*, 86(12):2218–2224, 1982.

[122] Bruce S Brunschwig, Jean Logan, Marshall D Newton, and Norman Sutin. A semiclassical treatment of electron-exchange reactions. application to the hexaaquoiron (ii)-hexaaquoiron (iii) system. *Journal of the American Chemical Society*, 102(18):5798–5809, 1980.

[123] LD Landau. Zur theorie der energieubertragung i. *Z. Sowjetunion*, 1:88–95, 1932.

[124] LD Landau. On the theory of transfer of energy at collisions ii. *Phys. Z. Sowjetunion*, 2(46):118, 1932.

[125] Clarence Zener. Non-adiabatic crossing of energy levels. *Proc. R. Soc. Lond. A*, 137(833):696–702, 1932.

[126] Jacob Spencer, Laura Scalfi, Antoine Carof, and Jochen Blumberger. Confronting surface hopping molecular dynamics with marcus theory for a molecular donor–acceptor system. *Faraday discussions*, 195:215–236, 2017.

[127] Rudolph A Marcus. On the theory of oxidation-reduction reactions involving electron transfer. i. *The Journal of Chemical Physics*, 24(5):966–978, 1956.

[128] Rudolph A Marcus. Electron transfer reactions in chemistry. theory and experiment. *Reviews of Modern Physics*, 65(3):599, 1993.

[129] R.A. Marcus and Norman Sutin. Electron transfers in chemistry and biology. *Biochimica et Biophysica Acta (BBA) - Reviews on Bioenergetics*, 811(3):265 – 322, 1985.

[130] Alexander Heck, Julian J. Kranz, Tomáš Kubař, and Marcus Elstner. Multi-Scale Approach to Non-Adiabatic Charge Transport in High-Mobility Organic Semiconductors. *Journal of Chemical Theory and Computation*, 11(11):5068–5082, November 2015.

[131] X. Jiang, Z. Futera, Md. E. Ali, F. Gajdos, G. F. von Rudorff, A. Carof, M. Breuer, and J. Blumberger. Cysteine linkages accelerate electron flow through tetra-heme protein STC. *J. Am. Chem. Soc.*, 139:17237–17240, 2017.

[132] Tomáš Kubař, P. Benjamin Woiczikowski, Gianaurelio Cuniberti, and Marcus Elstner. Efficient Calculation of Charge-Transfer Matrix Elements for Hole Transfer in DNA. *The Journal of Physical Chemistry B*, 112(26):7937–7947, 2008.

[133] Tomáš Kubař and Marcus Elstner. Efficient algorithms for the simulation of non-adiabatic electron transfer in complex molecular systems: application to DNA. *Physical Chemistry Chemical Physics*, 15(16):5794–5813, March 2013.

[134] Tomáš Kubař and Marcus Elstner. Coarse-Grained Time-Dependent Density Functional Simulation of Charge Transfer in Complex Systems: Application to Hole Transfer in DNA. *The Journal of Physical Chemistry B*, 114(34):11221–11240, September 2010.

[135] Tomáš Kubař and Marcus Elstner. What Governs the Charge Transfer in DNA? The Role of DNA Conformation and Environment. *The Journal of Physical Chemistry B*, 112(29):8788–8798, 2008.

[136] Alexander Heck, Julian J. Kranz, and Marcus Elstner. Simulation of Temperature-Dependent Charge Transport in Organic Semiconductors with Various Degrees of Disorder. *J. Chem. Theory Comput.*, 12(7):3087–3096, July 2016.

[137] Michael Mastalerz, Lucas Ueberricke, Daniel Holub, Julian Kranz, Frank Rominger, and Marcus Elstner. Triptycene end-capped quinoxalinophenanthrophenazines (qpps) – influence of substituents and conditions on aggregation in the solid state. *Chemistry – A European Journal*, 0(ja).

[138] Per-Olov Lowdin. On the Non-Orthogonality Problem Connected with the Use of Atomic Wave Functions in the Theory of Molecules and Crystals. *The Journal of Chemical Physics*, 18(3):365–375, March 1950.

[139] Yves A. Mantz, Francesco Luigi Gervasio, Teodoro Laino, and Michele Parrinello. Charge localization in stacked radical cation dna base pairs and the benzene dimer studied by self-interaction corrected density-functional theory. *The Journal of Physical Chemistry A*, 111(1):105–112, 2007.

[140] John C Tully. Molecular dynamics with electronic transitions. *The Journal of Chemical Physics*, 93(2):1061–1071, 1990.

[141] Giovanni Granucci and Maurizio Persico. Critical appraisal of the fewest switches algorithm for surface hopping. *The Journal of chemical physics*, 126(13):134114, 2007.

[142] Ahren W Jasper, Chaoyuan Zhu, Shikha Nangia, and Donald G Truhlar. Introductory lecture: Nonadiabatic effects in chemical dynamics. *Faraday Discussions*, 127:1–22, 2004.

[143] Linjun Wang and David Beljonne. Flexible surface hopping approach to model the crossover from hopping to band-like transport in organic crystals. *The journal of physical chemistry letters*, 4(11):1888–1894, 2013.

[144] Julian J Kranz and Marcus Elstner. Simulation of singlet exciton diffusion in bulk organic materials. *Journal of chemical theory and computation*, 12(9):4209–4221, 2016.

[145] William Humphrey, Andrew Dalke, and Klaus Schulten. VMD – Visual Molecular Dynamics. *Journal of Molecular Graphics*, 14:33–38, 1996.

[146] Melanie J Maul, Thomas RM Barends, Andreas F Glas, Max J Cryle, Tatiana Domratcheva, Sabine Schneider, Ilme Schlichting, and Thomas Carell. Crystal structure and mechanism of a dna (6-4) photolyase. *Angewandte Chemie International Edition*, 47(52):10076–10080, 2008.

[147] H. J. C. Berendsen, D. van der Spoel, and R. van Drunen. GROMACS: A message-passing parallel molecular dynamics implementation. *Computer Physics Communications*, 91(1):43–56, September 1995.

[148] Mark James Abraham, Teemu Murtola, Roland Schulz, Szilárd Páll, Jeremy C. Smith, Berk Hess, and Erik Lindahl. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1–2:19–25, September 2015.

[149] Junmei Wang, Piotr Cieplak, and Peter A. Kollman. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *Journal of Computational Chemistry*, 21(12):1049–1074, September 2000.

[150] Viktor Hornak, Robert Abel, Asim Okur, Bentley Strockbine, Adrian Roitberg, and Carlos Simmerling. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Structure, Function, and Bioinformatics*, 65(3):712–725, November 2006.

[151] Junmei Wang, Wei Wang, Peter A. Kollman, and David A. Case. Automatic atom type and bond type perception in molecular mechanical calculations. *Journal of Molecular Graphics and Modelling*, 25(2):247 – 260, 2006.

[152] Junmei Wang, Romain M. Wolf, James W. Caldwell, Peter A. Kollman, and David A. Case. Development and testing of a general amber force field. *Journal of Computational Chemistry*, 25(9):1157–1174, 2004.

[153] Brent H. Besler, Kenneth M. Merz, and Peter A. Kollman. Atomic charges derived from semiempirical methods. *Journal of Computational Chemistry*, 11(4):431–439, 1990.

[154] G. A. Petersson, Andrew Bennett, Thomas G. Tensfeldt, Mohammad A. Al-Laham, William A. Shirley, and John Mantzaris. A complete basis set model chemistry. I. The total energies of closed-shell atoms and hydrides of the first-row elements. *The Journal of Chemical Physics*, 89(4):2193–2218, 1988.

[155] Michael J. Frisch, G. W. Trucks, H. Bernhard Schlegel, Gustavo E. Scuseria, Michael A. Robb, James R. Cheeseman, Giovanni Scalmani, Vincenzo Barone, Benedetta Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, Xiaosong Li, H. P. Hratchian, Artur F. Izmaylov, Julien Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery Jr., J. E. Peralta, François Ogliaro, Michael J. Bearpark, Jochen Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, Rika Kobayashi, J. Normand, Krishnan Raghavachari, Alistair P. Rendell, J. C. Burant, S. S. Iyengar, Jacopo Tomasi, M. Cossi, N. Rega, N. J. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ödön Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and Douglas J. Fox. *Gaussian 09*. Gaussian, Inc., Wallingford, CT, USA, 2009.

[156] Markus Kaukonen, Pär Söderhjelm, Jimmy Heimdal, and Ulf Ryde. Proton transfer at metal sites in proteins studied by quantum mechanical free-energy perturbations. *Journal of Chemical Theory and Computation*, 4(6):985–1001, 2008.

[157] Christopher H. Chang and Kwiseon Kim. Density functional theory calculation of bonding and charge parameters for molecular dynamics studies on [fefe] hydrogenases. *Journal of Chemical Theory and Computation*, 5(4):1137–1145, 2009.

[158] Andrej Šali and Tom L. Blundell. Comparative protein modelling by satisfaction of spatial restraints. *Journal of Molecular Biology*, 234(3):779 – 815, 1993.

[159] William L. Jorgensen, Jayaraman Chandrasekhar, Jeffry D. Madura, Roger W. Impey, and Michael L. Klein. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, 79(2):926–935, 1983.

[160] Berk Hess, Henk Bekker, Herman J. C. Berendsen, and Johannes G. E. M. Fraaije. LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry*, 18(12):1463–1472, September 1997.

[161] Martin J. Field, Paul A. Bash, and Martin Karplus. A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations. *Journal of Computational Chemistry*, 11(6):700–733, 1990.

[162] DN Beratan, JN Betts, and JN Onuchic. Protein electron transfer rates set by the bridging secondary and tertiary structure. *Science*, 252(5010):1285–1288, 1991.

[163] R. Al Marcus and Norman Sutin. Electron transfers in chemistry and biology. *Biochimica et Biophysica Acta (BBA)-Reviews on Bioenergetics*, 811(3):265–322, 1985.

[164] Jeng-Da Chai and Martin Head-Gordon. Long-range corrected hybrid density functionals with damped atom–atom dispersion corrections. 10(44):6615–6620, November 2008.

[165] G. A. Petersson and Mohammad A. Al-Laham. A complete basis set model chemistry. II. Open-shell systems and the total energies of the first-row atoms. *The Journal of Chemical Physics*, 94(9):6081–6090, 1991.

[166] Chaitanya Saxena, Aziz Sancar, and Dongping Zhong. Femtosecond dynamics of dna photolyase energy transfer of antenna initiation and electron transfer of cofactor reduction. *The Journal of Physical Chemistry B*, 108(46):18026–18033, 2004.

[167] Mario Wolter, Marcus Elstner, Ulrich Kleinekathöfer, and Tomáš Kubař. Microsecond Simulation of Electron Transfer in DNA: Bottom-Up Parametrization of an Efficient Electron Transfer Model Based on Atomistic Details. *The Journal of Physical Chemistry B*, 121(3):529–549, January 2017.

[168] Adam Kubas, Felix Hoffmann, Alexander Heck, Harald Oberhofer, Marcus Elstner, and Jochen Blumberger. Electronic couplings for molecular charge transfer: Benchmarking CDFT, FODFT, and FODFTB against high-level ab initio calculations. *The Journal of Chemical Physics*, 140(10):104105–21, March 2014.

[169] Yann Geisselbrecht, Sebastian Frühwirth, Claudia Schroeder, Antonio J Pierik, Gabriele Klug, and Lars-Oliver Essen. Cryb from rhodobacter sphaeroides: a unique class of cryptochromes with new cofactors. *EMBO reports*, 13(3):223–229, 2012.

[170] Hisham M Dokainish and Akio Kitao. Computational assignment of the histidine protonation state in (6-4) photolyase enzyme and its effect on the protonation step. *ACS Catalysis*, 6(8):5500–5507, 2016.

[171] Jiang Li, Zheyun Liu, Chuang Tan, Xunmin Guo, Lijuan Wang, Aziz Sancar, and Dongping Zhong. Dynamics and mechanism of repair of ultraviolet-induced (6–4) photoproduct by photolyase. *Nature*, 466(7308):887, 2010.

[172] Zheyun Liu, Chuang Tan, Xunmin Guo, Ya-Ting Kao, Jiang Li, Lijuan Wang, Aziz Sancar, and Dongping Zhong. Dynamics and mechanism of cyclobutane pyrimidine dimer repair by dna photolyase. *Proceedings of the National Academy of Sciences*, 108(36):14831–14836, 2011.

[173] Shirin Faraji, Dongping Zhong, and Andreas Dreuw. Characterization of the intermediate in and identification of the repair mechanism of (6-4) photolesions by photolyases. *Angewandte Chemie*, 128(17):5261–5264, 2016.

[174] Dongping Zhong. Electron transfer mechanisms of dna repair by photolyase. *Annual review of physical chemistry*, 66:691–715, 2015.

[175] Shirin Faraji and Andreas Dreuw. Insights into light-driven dna repair by photolyases: Challenges and opportunities for electronic structure theory. *Photochemistry and photobiology*, 93(1):37–50, 2017.

[176] Ilia A Solov'yov, Andrey V Korol, and Andrey V Solov'yov. *Multiscale modeling of complex molecular structure and dynamics with MBN explorer*. Springer, 2017.

[177] Yuma Terai, Ryuma Sato, Takahiro Yumiba, Ryuhei Harada, Kohei Shimizu, Tatsuya Toga, Tomoko Ishikawa-Fujiwara, Takeshi Todo, Shigenori Iwai, Yasuteru Shigeta, et al. Coulomb and ch–π interactions in (6–4) photolyase–dna complex dominate dna binding and repair abilities. *Nucleic acids research*, 2018.

[178] Katrine Aalbæk Jepsen and Ilia A Solov'yov. On binding specificity of (6–4) photolyase to a t (6–4) t dna photoproduct. *The European Physical Journal D*, 71(6):155, 2017.

[179] Andrea Christa Kneuttinger, Gengo Kashiwazaki, Stefan Prill, Korbinian Heil, Markus Müller, and Thomas Carell. Formation and direct repair of uv-induced dimeric dna pyrimidine lesions. *Photochemistry and photobiology*, 90(1):1–14, 2014.

[180] Junpei Yamamoto, Pascal Plaza, and Klaus Brettel. Repair of (6-4) lesions in dna by (6-4) photolyase: 20 years of quest for the photoreaction mechanism. *Photochemistry and photobiology*, 93(1):51–66, 2017.

[181] Fabien Cailliez, Pavel Müller, Thiago Firmino, Pascal Pernot, and Aurélien de la Lande. Energetics of Photoinduced Charge Migration within the Tryptophan Tetrad of an Animal (6-4) Photolyase. *Journal of the American Chemical Society*, 138:1904, 2016.

[182] Daniel Holub, Tilman Lamparter, Marcus Elstner, and Natacha Gillet. Biological relevance of charge transfer branching pathways in photolyases. *Physical Chemistry Chemical Physics*, 2019 (submitted).

[183] Pavel Müller, Junpei Yamamoto, Ryan Martin, Shigenori Iwai, and Klaus Brettel. Discovery and functional analysis of a 4th electron-transferring tryptophan conserved exclusively in animal cryptochromes and (6-4) photolyases. *Chem. Commun.*, 51:15502–15505, 2015.

[184] Junpei Yamamoto, Kohei Shimizu, Takahiro Kanda, Yuhei Hosokawa, Shigenori Iwai, Pascal Plaza, and Pavel Müller. Loss of fourth electron-transferring tryptophan in animal (6–4) photolyase impairs dna repair activity in bacterial cells. *Biochemistry*, 56(40):5356–5364, 2017.

[185] Ilia A. Solov'yov, Tatiana Domratcheva, Abdul Rehaman Moughal Shahi, and Klaus Schulten. Decrypting Cryptochrome: Revealing the Molecular Identity of the Photoactivation Reaction. *Journal of the American Chemical Society*, 134(43):18046–18052, October 2012.

[186] Thiago Firmino, Etienne Mangaud, Fabien Cailliez, Adrien Devolder, David Mendive-Tapia, Fabien Gatti, Christoph Meier, Michele Desouter-Lecomte, and Aurelien de la Lande. Quantum effects in ultrafast electron transfers within cryptochromes. *Phys. Chem. Chem. Phys.*, 18:21442–21457, 2016.

[187] Ryan Martin, Fabien Lacombat, Agathe Espagne, Nadia Dozova, Pascal Plaza, Junpei Yamamoto, Pavel Muller, Klaus Brettel, and Aurelien de la Lande. Ultrafast flavin photoreduction in an oxidized animal (6-4) photolyase through an unconventional tryptophan tetrad. *Phys. Chem. Chem. Phys.*, 19:24493–24504, 2017.

[188] T. Kubar and M. Elstner. A hybrid approach to simulation of electron transfer in complex molecular systems. *Journal of The Royal Society Interface*, 10(87):20130415–20130415, July 2013.

[189] Asif Bashir, Alexander Heck, Akimitsu Narita, Xinliang Feng, Alexei Nefedov, Michael Rohwerder, Klaus Mullen, Marcus Elstner, and Christof Woll. Charge carrier mobilities in organic semiconductors: crystal engineering and the importance of molecular contacts. *Phys. Chem. Chem. Phys.*, 17:21988–21996, 2015.

[190] David E. Shaw, Paul Maragakis, Kresten Lindorff-Larsen, Stefano Piana, Ron O. Dror, Michael P. Eastwood, Joseph A. Bank, John M. Jumper, John K. Salmon, Yibing Shan, and Willy Wriggers. Atomic-level characterization of the structural dynamics of proteins. *Science*, 330(6002):341–346, 2010.

[191] U. Chandra Singh and Peter A. Kollman. An approach to computing electrostatic charges for molecules. *Journal of Computational Chemistry*, 5(2):129–145, 1984.

[192] Chresten R. Søndergaard, Mats H. M. Olsson, Michał Rostkowski, and Jan H. Jensen. Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pKa Values. *Journal of Chemical Theory and Computation*, 7(7):2284–2295, 2011.

[193] Mats H. M. Olsson, Chresten R. Søndergaard, Michal Rostkowski, and Jan H. Jensen. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *Journal of Chemical Theory and Computation*, 7(2):525–537, 2011.

[194] Pavel Müller, Elisabeth Ignatz, Stephan Kiontke, Klaus Brettel, and Lars-Oliver Essen. Sub-nanosecond tryptophan radical deprotonation mediated by a protein-bound water cluster in class ii dna photolyases. 9:1200–1212, 12 2017.

[195] Pavel Müller, Jean-Pierre Bouly, Kenichi Hitomi, Véronique Balland, Elizabeth D Getzoff, Thorsten Ritz, and Klaus Brettel. Atp binding turns plant cryptochrome into an efficient natural photoswitch. 4:5175, 06 2014.

[196] Corinne Aubert, Marten H Vos, Paul Mathis, André Eker, and Klaus Brettel. Intraprotein radical transfer during photoactivation of dna photolyase. *Nature*, 405:586–590, 2000.

[197] Ywan Feng Li, Paul F. Heelis, and Aziz Sancar. Active site of dna photolyase: tryptophan-306 is the intrinsic hydrogen atom donor essential for flavin radical photoreduction and dna repair in vitro. *Biochemistry*, 30(25):6322–6329, 1991.

[198] Johanna Brazard, Anwar Usman, Fabien Lacombat, Christian Ley, Monique M. Martin, and Pascal Plaza. New insights into the ultrafast photophysics of oxidized and reduced fad in solution. *The Journal of Physical Chemistry A*, 115(15):3251–3262, 2011.

[199] A. Carof, S. Giannini, and J. Blumberger. Detailed balance, internal consistency and energy conservation in fragment orbital-based surface hopping. *J. Chem. Phys.*, 147:214113, 2017.

[200] S. Giannini, A. Carof, and J. Blumberger. Crossover from hopping to band-like charge transport in an organic semiconductor model: Atomistic non-adiabatic molecular dynamics simulation . *J. Phys. Chem. Lett*, 9:3116–3123, 2018.

[201] RoGo Kepler. Charge carrier production and mobility in anthracene crystals. *Physical Review*, 119(4):1226, 1960.

[202] Yusei Maruyama and Hiroo Inokuchi. Charge-carrier mobility in anthracene single crystals. *Bulletin of the Chemical Society of Japan*, 40(9):2073–2077, 1967.

[203] C Pratt Brock and JD Dunitz. Temperature dependence of thermal motion in crystalline anthracene. *Acta Crystallographica Section B: Structural Science*, 46(6):795–806, 1990.

[204] Th Holstein. Studies of polaron motion: Part i. the molecular-crystal model. *Annals of physics*, 8(3):325–342, 1959.

[205] Veaceslav Coropceanu, Jérôme Cornil, Demetrio A. da Silva Filho, Yoann Olivier, Robert Silbey, and Jean-Luc Brédas. Charge transport in organic semiconductors. *Chemical Reviews*, 107(4):926–952, 2007.

[206] Jenny Nelson, Joe J. Kwiatkowski, James Kirkpatrick, and Jarvist M. Frost. Modeling charge transport in organic photovoltaic materials. *Accounts of Chemical Research*, 42(11):1768–1778, 2009.

[207] Alexey V Akimov, Run Long, and Oleg V Prezhdo. Coherence penalty functional: A simple method for adding decoherence in ehrenfest dynamics. *The Journal of chemical physics*, 140(19):194107, 2014.

[208] Norbert Karl and Jörg Marktanner. Electron and hole mobilities in high purity anthracene single crystals. *Molecular Crystals and Liquid Crystals Science and Technology. Section A. Molecular Crystals and Liquid Crystals*, 355(1):149–173, 2001.

[209] Daniel Holub, Tomáš Kubař, Thilo Mast, Marcus Elstner, and Natacha Gillet. What accounts for the different functions in photolyases and cryptochromes: a computational study of proton transfers to fad. *Physical Chemistry Chemical Physics*, 2019.

[210] Michael J. Damiani, Jordan J. Nostedt, and Melanie A. O'Neill. Impact of the N5-proximal Asn on the Thermodynamic and Kinetic Stability of the Semiquinone Radical in Photolyase. *Journal of Biological Chemistry*, 286(6):4382–4391, February 2011.

[211] Chad A. Brautigam, Barbara S. Smith, Zhiquan Ma, Maya Palnitkar, Diana R. Tomchick, Mischa Machius, and Johann Deisenhofer. Structure of the photolyase-like domain of cryptochrome 1 from arabidopsis thaliana. *Proceedings of the National Academy of Sciences*, 101(33):12142–12147, 2004.

[212] 2014. AmberTools 14, available at: http://ambermd.org. AmberTools 14, available at: http://ambermd.org, 2014.

[213] Shankar Kumar, John M. Rosenberg, Djamal Bouzida, Robert H. Swendsen, and Peter A. Kollman. The weighted histogram analysis method for free-energy calculations on biomolecules. i. the method. *Journal of Computational Chemistry*, 13(8):1011–1021, 1992.

[214] A. grossfield, "wham: the weighted histogram analysis method", version 2.0.9, http://membrane.urmc.rochester.edu/content/wham, .

[215] Massimiliano Bonomi, Davide Branduardi, Giovanni Bussi, Carlo Camilloni, Davide Provasi, Paolo Raiteri, Davide Donadio, Fabrizio Marinelli, Fabio Pietrucci, Ricardo A. Broglia, and Michele Parrinello. Plumed: A portable plugin for free-energy calculations with molecular dynamics. *Computer Physics Communications*, 180(10):1961 – 1972, 2009.

[216] Gareth A. Tribello, Massimiliano Bonomi, Davide Branduardi, Carlo Camilloni, and Giovanni Bussi. Plumed 2: New feathers for an old bird. *Computer Physics Communications*, 185(2):604 – 613, 2014.

[217] P. H. König, N. Ghosh, M. Hoffmann, M. Elstner, E. Tajkhorshid, Th. Frauenheim, and Q. Cui. Toward theoretical analyis of long-range proton transfer kinetics in biomolecular pumps. *The Journal of Physical Chemistry A*, 110(2):548–563, 2006.

[218] Carrie L. Partch, Michael W. Clarkson, Sezgin Özgür, Andrew L. Lee, and Aziz Sancar. Role of structural plasticity in signal transduction by the cryptochrome blue-light photoreceptor. *Biochemistry*, 44(10):3795–3805, 2005.

[219] Ya-Ting Kao, Chuang Tan, Sang-Hun Song, Nuri Öztürk, Jiang Li, Lijuan Wang, Aziz Sancar, and Dongping Zhong. Ultrafast Dynamics and Anionic Active States of the Flavin Cofactor in Cryptochrome and Photolyase. *Journal of the American Chemical Society*, 130(24):7695–7701, June 2008.

[220] Karl F. Herzfeld. Zur Theorie der Reaktionsgeschwindigkeiten in Gasen. *Annalen der Physik*, 364(15):635–667.

[221] Bernd Paulus, Csaba Bajzath, Frédéric Melin, Lorenz Heidinger, Viktoria Kromm, Christoph Herkersdorf, Ulrike Benz, Lisa Mann, Patricia Stehle, Petra Hellwig, Stefan Weber, and Erik Schleicher. Spectroscopic characterization of radicals and radical pairs in fruit fly cryptochrome – protonated and nonprotonated flavin radical-states. *FEBS J*, 282(16):3175–3189, August 2015.

Publications

12/2017	Functional role of an unusual tyrosine residue in the electron transfer chain of a prokaryotic (6–4) photolyase, Holub, D., Ma, H., Krauß, N., Lamparter, T., Elstner, M., Gillet, N. *Chem. Sci*. 2018
01/2019	Two aspartate residues close to the lesion binding site of agrobacterium (6-4) photolyase are required for Mg^{2+} stimulation of DNA repair, M. Hongju, Holub D., Gillet N., Kaeser G., Thoulass K., Elstner M., Krausß N. and Lamparter T., *The FEBS Journal*, 2019
02/2019	What accounts for the different functions in photolyases and cryptochromes: a computational study of proton transfers to FAD, Holub D., Kubař T., Mast T., Elstner M., and Gillet N., Phys. Chem. Chem. Phys., 2019
03/2019	Biological relevance of charge transfer branching pathways in photolyases, Holub D., Lamparter T., Elstner M., and Gillet N., Physical Chemistry Chemical Physics, 2019 (*submitted*).
03/2019	Triptycene End-capped Quinoxalinophenanthrophenazines (QPP) – Influence of Substituents and Conditions on Aggregation in the Solid State Ueberricke L., Holub D., Kranz J., Rominger F., Elstner M. and Mastalerz M. Chemistry – A European Journal, 2019

Acknowledgement

First of all, I would like to express my very great appreciation to my adviser Prof. Dr. Marcus Elstner for giving me the opportunity to work on this project. He was always available for discussions and offered scientific advice, furthermore, he gave me the freedom to expand my research according to my own interest. I would also like to thank Prof. Dr. Wolfgang Wenzel for being the co-referee of this thesis.

Furthermore, thank you to Natacha Gillet and Tomáš Kubař, two post-doctors of the group who supported me during the last years. Firstly, Tomáš developed the method, which I used for all the charge transfer simulations. Therefore, I learned much from him, which gave me the knowledge to extend the method.
Natacha helped me to clarify the questions of my different biological systems, thus, we spent many hours discussing to relate the simulated results with the experimental observations. We also travelled often to participate at conferences and for our collaborations.

Additionally, I would like to thank the whole TCB group. They are the best colleagues and we had a lot of productive coffee breaks and enjoyed many barbecues at the rooftop of the Girls-Tower.

Besides, I would like to thank the KHYS for the Travel-Grand, to support my exchange to London. In addition to that, I also want to thank Prof. Dr. Jochen Blumberger for the supervision in London and his whole group.

Last but not least, a very special thanks to my family and Ilona Wehl. Without their help, support, encouragement and motivation during the last years this wouldn't be possible.