Multiscale Simulations of Proton Coupled Electron Transfer (PCET) in Biological Systems

Zur Erlangung des akademischen Grades einer

DOKTORIN DER NATURWISSENSCHAFTEN

(Dr. rer. nat.)

von der KIT-Fakultät für Chemie und Biowissenschaften
des Karlsruher Instituts für Technologie (KIT)

genehmigte
DISSERTATION

von
M.Sc. Katharina Spies
aus Wiesbaden

1. Referent:	Prof. Marcus Elstner
2. Referentin:	Dr. Natacha Gillet
Tag der mündlichen Prüfung:	21.10.2025

ENS DE LYON

Karlsruher Institut für Technologie

THESE

pour l'obtention du grade de Docteur, délivré par

l'ECOLE NORMALE SUPERIEURE DE LYON

en cotutelle avec

KARLSRUHE INSTITUTE OF TECHNOLOGY

Ecole Doctorale N°206
Chimie de Lyon

Discipline : CHIMIE (Chimie)

Soutenue publiquement le 21 octobre 2025, par :

Katharina SPIES

Simulations multi-échelles du transfert d'électrons couplé à des protons (PCET) dans les systèmes biologiques

Multiscale Simulations of Proton Coupled Electron Transfer (PCET) in Biological Systems

---

Après avis de :

Elise DUMONT, Professeure des universités, Université Côte d'Azur	Rapporteure
Adèle LAURENT, Directrice de recherche, Université de Nantes / CNRS	Rapporteure

Devant le jury composé de :

Elise DUMONT, Professeure des universités, Université Côte d'Azur	Rapporteure
Adèle LAURENT, Directrice de recherche, Université de Nantes / CNRS	Rapporteure
Sebastian HÖFENER, Chercheur, Institut de Technologie de Karlsruhe (KIT)	Examinateur
Rebecca WADE, Professeure, Université d'Heidelberg	Examinatrice
Carine MICHEL, Directrice de recherche, ENS de Lyon / CNRS	Examinatrice
Natacha GILLET, Chargée de recherche - HDR, ENS de Lyon / CNRS	Directrice de thèse
Marcus ELSTNER, Professeur, Institut de technologie de Karlsruhe	Co-directeur de thèse

Abstract

Computational chemistry has demonstrated its ability to provide insightful mechanical data in the study of biomolecules. However, it still faces challenges in describing the complexity of these systems and overcoming the limits due to their size. QM/MM multiscale approaches, combining quantum mechanical descriptions of chemical reactions with classical force fields for the environment, are powerful tools for protein chemical properties and reactivity, especially when coupled with enhanced sampling methods.

In this PhD work, these methods are applied to proton-coupled electron transfer (PCET) reactions, crucial in systems ranging from photosynthesis to radical catalysis. We developed a QM/MM protocol combined with metadynamics simulations and applied it to biomimetic peptides. To reduce computing cost, the QM region is treated using the semi-empirical DFTB3 method. The proton transfer collective variable (CV) was biased during metadynamics simulations, while the electron transfer CV was analyzed in post-processing. Free energy surfaces yielded mechanistic and thermodynamic insights such as reaction free energy barriers, enabling the systematic investigation of the influence of different PCET partners (radical tyrosine with histidine, tryptophan, or tyrosine), different configurations, solvent exposure, and protein embedding.

Thanks to efficient sampling, our protocol was applied to proteins. Photolyases and cryptochromes share a photoactivation process where a flavin derivative is reduced by a tryptophan triad. Experiments suggest that in class II photolyase from archaea *Methanosarcina mazei* (MmCPDII) and in animal-like cryptochrome from green alga *Chlamydomonas reinhardtii* (CraCRY), a tyrosine acts as the terminal electron donor. Our method identified a multiple-partner PCET mechanism in MmCPDII where tryptophan accepts the electron and histidine the proton, while in cryptochrome, water-mediated deprotonation followed by PCET between the tryptophan radical and tyrosine is likely. We successfully demonstrate that the close residues impact the nature of the PCET mechanism.

Despite the protocol's advantages, limitations remain in the description of the electron transfer CV. To address this, we applied a bias potential to the electron transfer using a DFTB version in which coupled-perturbed (CP) equations are implemented. As solving these equations is computationally costly, we demonstrate that restricting the calculations to a subset of MM atoms via a cutoff reduces cost, thereby enabling application to the biomimetic peptides. This development paves the way to a more consistent protocol, allowing a better description of the electron behavior.

In addition, QM/MM-based simulations were also used to study the spectroscopic properties of retinylidene proteins. We present a machine learning approach where retinal chromophores (11-cis and all-trans) were simulated in various solvents and ion environments to generate diverse electrostatic conditions mimicking proteins. Excited state calculations on simulation snapshots using different QM methods provided training data for a neural network, which then predicted absorption spectra from protein simulations.

Finally, we focus on the channelrhodopsin (ChR) Chrimson, highlighting how active site structure, counterion arrangement, and hydrogen bonding networks tune retinal absorption. Hybrid quantum mechanics/molecular mechanics (QM/MM) simulations of models with varying protonation states of titrable residues, followed by excited-state calculations, reveal a highly flexible active site. We identify that hydrogen bonding between counterions contributes to Chrimson's uniquely red-shifted absorption.

Résumé

La chimie computationnelle s'est révélée une méthode de choix pour l'étude des mécanismes chimiques à l'œuvre dans les biomolécules. Cependant, des défis restent encore à relever pour décrire la complexité de ces systèmes et dépasser les limites associées à leur taille. Les approches multi-échelles QM/MM (combinant des descriptions au niveau quantique des réactions chimiques avec des champs de force classiques pour l'environnement), alliées à des méthodes d'échantillonnage renforcé, s'avèrent être des outils pertinents pour la détermination des propriétés chimiques et de la réactivité des protéines.

Dans ce travail de doctorat, ces méthodes sont appliquées aux réactions de transferts couplés d'électron-proton (PCET), cruciales dans des systèmes allant de la photosynthèse à la catalyse des radicaux. Nous avons mis au point un protocole QM/MM combiné à des simulations de métadynamique et l'avons appliqué à des peptides biomimétiques. Pour réduire les coûts de calcul, la région QM est traitée au niveau semi-empirique DFTB3. La variable collective (CV) de transfert de proton est biaisée, tandis que la CV de transfert d'électron est analysée en post-traitement. Les surfaces d'énergie libre ont fourni des informations mécanistiques et thermodynamiques telles que les barrières d'énergie libre de réaction, permettant ainsi une étude systématique de l'influence de différents partenaires impliqués, de leurs différentes configurations, de l'exposition au solvant ou de l'environnement protéique.

Grâce à un échantillonnage efficace, notre protocole a pu être appliqué à d'autres protéines. Les photolyases et les cryptochromes partagent un processus de photoactivation dans lequel un dérivé de flavine est réduit par une triade de tryptophanes. Des expériences suggèrent que dans la photolyase de classe II de *Methanosarcina mazei* (MmCPDII) et le cryptochrome de type animal de *Chlamydomonas reinhardtii* (CraCRY), une tyrosine joue le rôle de donneur d'électron terminal. Notre méthode a permis d'identifier un mécanisme de PCET à partenaires multiples dans le MmCPDII, tandis que dans le cryptochrome, une déprotonation via une chaîne de molécules d'eau suivie d'un PCET entre le tryptophane radicalaire et la tyrosine est probable. Nous démontrons donc avec succès que les acides aminés proches ont un impact sur la nature du mécanisme de PCET.

Malgré les avantages de ce protocole, des limitations subsistent dans la description du transfert d'électron. Pour y remédier, nous avons appliqué un potentiel de biais au transfert d'électron en utilisant une version de la méthode DFTB dans laquelle les équations à perturbation couplée (CP) sont implémentées. La résolution de ces équations étant coûteuse en termes de calcul mais nous démontrons que la restriction des calculs à un sous-ensemble d'atomes MM réduit le coût, ce qui permet l'application aux peptides biomimétiques. Ce développement ouvre la voie à un protocole permettant une meilleure description du comportement des électrons.

Des simulations basées sur QM/MM ont également été utilisées pour étudier les propriétés spectroscopiques des protéines rétinylidènes. Nous présentons une approche d'apprentissage automatique dans laquelle des chromophores rétiniens (11-cis et tout-trans) ont été simulés dans différents solvants et environnements ioniques afin de générer diverses conditions électrostatiques imitant les protéines. Les calculs de l'état excité sur des instantanés de simulation utilisant différentes méthodes QM ont fourni des données d'entraînement pour un réseau neuronal, qui a ensuite prédit des spectres d'absorption à partir de simulations de protéines.

Enfin, nous nous concentrons sur la channelrhodopsine Chrimson. Les simulations QM/MM de modèles avec différents états de protonation des résidus titrables, suivies de modélisations de l'état excité, révèlent un site actif très flexible. La liaison hydrogène entre les contre-ions contribue à l'absorption décalée vers le rouge observée uniquement dans cette protéine.

Zusammenfassung

Die computergestützte Chemie hat sich als wertvolles Instrument erwiesen, um aufschlussreiche mechanistische Einblicke in Biomolekülen zu gewinnen. Dennoch bestehen weiterhin Herausforderungen, die Komplexität dieser Systeme adäquat zu beschreiben und die durch ihre Systemgröße gesetzten Grenzen zu überwinden. QM/MM-Multiskalenansätze, bei denen chemische Reaktionen quantenmechanisch beschrieben werden und die Umgebung mit klassischen Kraftfeldern modelliert wird, bieten hierfür ein besonders leistungsfähiges Werkzeug, insbesondere in Kombination mit Methoden, die eine umfassende Erkundung der Konformationslandschaft ermöglichen.

In dieser Dissertation werden Protonen-gekoppelte Elektronentransferreaktionen (PCET) untersucht, die von zentraler Bedeutung für Prozesse wie Photosynthese und Radikalkatalyse sind. Dafür wurde ein QM/MM-Protokoll in Kombination mit Metadynamik entwickelt und auf biomimetische Peptide angewendet. Um den Rechenaufwand zu reduzieren, kam für die QM-Region die semi-empirische DFTB3-Methode zum Einsatz. Während der Simulation wurde die Reaktionskoordiante des Protonentransfers gezielt beeinflusst, um die Konformationslandschaft besser zu erkunden. Der Elektronentransfer wurde jedoch erst in der Nachbearbeitung analysiert. Die resultierenden freien Energieflächen liefern mechanistische und thermodynamische Einblicke, darunter Reaktionsbarrieren. Dies ermöglicht eine systematisch Analyse des Einflusses verschiedener PCET-Partner (radikales Tyrosin mit Histidin, Tryptophan oder Tyrosin), unterschiedlicher Konfigurationen und des Einflusses des Lösungsmittels oder der Proteinumgebung.

Dank Methoden wie Metadynamik, die die Konformationslandschaft effizient erkunden, konnte das Protokoll auch auf Proteine angewendet werden. Photolyasen und Cryptochrome durchlaufen einen ähnlichen Photoaktivierungsprozess, bei dem eine Tryptophan-Triade ein Flavin-Derivat reduziert. Experimentelle Befunde deuten darauf hin, dass in der Klasse-II-Photolyase aus der Archae *Methanosarcina mazei* (Mm-CPDII) und im tierähnlichen Cryptochrome der Alge *Chlamydomonas reinhardtii* (CraCRY) ein Tyrosin-Rest die Rolle als terminalen Elektronendonors übernimmt. Unsere Methode identifiziert in MmCPDII einen PCET-Mechanismus mit mehreren Partnern, bei dem das Elektron vom Tryptophan und das Proton vom Histidin aufgenommen wird. In CraCRY hingegen spricht vieles für eine Deprotonierung in Wasser, das in naher Umgebung ist, gefolgt von einem PCET zwischen dem Tryptophanradikal und Tyrosin. Es wird gezeigt, dass benachbarte Aminosäurereste die Art des PCET-Mechanismus entscheidend beeinflussen.

Trotz der Vorteile des Protokolls bestehen weiterhin Einschränkungen bei der Beschreibung des Elektronentransfers. Um dem entgegenzuwirken, wird auch die Reaktionskoordinate des Elektronentransfers beeinflusst. Das wird ermöglicht durch die Nutzung einer DFTB Version, bei der gekoppelte Störungsgleichungen gelöst werden. Da die Lösung dieser Gleichungen rechenintensiv ist, wird gezeigt, dass eine Beschränkung der Berechnungen auf eine reduzierte Anzahl an MM-Atomen den Rechenaufwand deutlich reduzieren kann. Somit wird die Anwendung der Methode auf biomimetische Peptide ermöglicht. Diese Entwicklung ebnet den Weg für ein konsistenteres Protokoll mit einer verbesserten Beschreibung des Elektronenverhaltens.

Zusätzlich wurden QM/MM-basierte Simulationen genutzt, um spektroskopische Eigenschaften von Retinalproteinen zu untersuchen. Es wird eine Methode vorgestellt, bei der maschinelles Lernen zum Einsatz kommt. Der Chromophor Retinal wird in seinen zwei Isomerformen (11-cis und all-trans) in verschiedenen Lösungsmitteln und Ionenumgebungen simuliert. Dies ist eine Nachahmung der vielfältigen elektrostatischen Bedingungen, wie sie in Proteinen vorkommen. Mithilfe von verschiedenen QM-Methoden werden spektroskopische Eigenschaften berechnet, die als Trainingsdaten für neuronale Netze dienen. Diese sagen anschließend Absorptionsspektren von Retinalproteinen voraus.

Abschließend wird eine Studie zu dem Kanalrhodopsin Chrimson durchgeführt. Dabei wird aufgezeigt, wie die Absorption von Retinal durch die Struktur des aktiven Zentrums, der Anordnung der Gegenionen und des Wasserstoffbrückenbindungsnetzwerkes beeinflusst wird. QM/MM-Simulationen von Modellen mit unterschiedlichen Protonierungszuständen titrierbarer Reste zeigen ein hochflexibles aktives Zentrum. Es zeigt sich durch Berechnungen im angeregten Zustand, dass eine Wasserstoffbrückenbindung zwischen den Gegenionen zur einzigartigen Rotverschiebung der Absorption von Chrimson beiträgt.

Acknowledgment

First of all, I would like to express my sincere gratitude to Prof. Marcus Elstner, not only for giving me the opportunity to pursue my PhD in his group, but especially for initiating the idea of a binational Co-Tutelle. This unique arrangement allowed me to spend three years between two groups in two amazing cities, and gave me the chance to grow far beyond scientific research alone. I am deeply grateful to Dr. Natacha Gillet for her outstanding supervision as my thesis director at the ENS de Lyon. You always knew exactly how to support me, and I particularly value your constructive criticism, which over the years has significantly improved my presentation and writing skills. Above all, you are an inspiring role model as a woman in science. I also want to thank Dr. Tomáš Kubař. Whether it was IT questions, thermodynamics, or just life in general, you always have an open ear and were ready to help. My thanks also go to Dr. Carine Michel, as director of the laboratory at the ENS de Lyon, for always competent and kind feedback and support in bureaucratic challenges. I do not want to forget Sabine Holthoff and Edwige Royboz for their continuous support, even when my situation was often unusual and more complicated.

I would like to thank everybody at the ENS lab for the amazing friendship and support, especially Laleh and Maxime. A big thanks also to Qing, the funniest office mate and to all the others: I truly love you guys <3.

Speaking about labs, I also thank the entire TCB group for four unforgettable years: from Indian lunches at Campus West (thanks to Deepak), to the best barbecues, hikes, and lake visits, to funny soccer-tournaments and Monday runs, and a great trip to Prague. Most of all, thank you for always being there for each other.

I do not want to forget Bea, who even years after leaving the group never stopped asking about the project's progress and would drop by for a nice cup of coffee and a chat. A big thanks goes to Lena for her detailed proofreading, constant ideas for figure improvements, and support with any question I brought up, I feel so lucky to have shared an office with you. Thanks to Julian as well: discussions about enhanced sampling, or any small details; you are always helpful and I truly appreciate your feedback. Lukas, you did an amazing proofreading-job, and you were always the first person I could turn to with coding problems. Without you, I would have been stuck countless times. To my students, Leonie, Béryl, Lara and Jemina: Thank you for advancing my research during your theses, helping me see my projects from new perspectives, and giving me the chance to gain experience in project management and leadership. I also want to thank David, Eik, and Maike, our discussions were always helpful, and I'm grateful for your input.

My friends have been a constant source of support. First, my roommate Astrid, who has been with me through thick and thin for almost ten years, throughout my entire studies and PhD. I am so grateful and honored to have you as a friend, I truly don't know how to thank you for everything. To all others, especially Annika, Laura, Marta, and Svenja who kept me going when things got tough, offering constant support.

To my boyfriend, Sergio, who had to deal with a nervous wrack during the last weeks, spoiled me when necessary, and visited me wherever I was at the time (thanks for loving Lyon almost as much as I do), thank you for being there.

To my sister Laura and her partner, Jorge, who sent the much-needed fuel in form of take-away meals and were only a phone call away whenever I needed an honest opinion. To my parents, who have supported me in everything since always, what else can I say than thank you! To the rest of my family, who will never understand what I'm actually working on but support me anyways as much as they can.

Finally, the biggest thanks to everybody: I truly appreciate each and every one of you, including those I haven't mentioned by name but who were there for me along the way.

Contents

Acronyms

ACE acetyl

ATP adenosine triphosphate

BLA bond length alternation

BestR bestrhodopsin

BO Born-Oppenheimer

bR bacteriorhodopsin

C cystein

ChR channelrhodopsin

ChR2 channelrhodopsin 2

CI configuration interaction

CraCRY animal-like cryptochrome from green alga *Chlamydomonas reinhardtii*

COM center of mass

CpHMD constant pH molecular dynamics

CPD cyclobutane pyrimidine dimers

CV collective variable

CP coupled-perturbed

D aspartate

DFT density functional theory

DFTB density functional tight binding

DNA deoxyribonucleic acid

DSSP Deffine Secondary Structure of Proteins

E glutamate

e.g. for example

EPT electron proton transfer

ESP electrostatic potential

ET electron transfer

FAD flavine adenine dinucleotide

FES free energy surface

GGA Generalized Gradient Approximation

GTOs Gaussian-type Orbitals

H histidine

HAT hydrogen atom transfer

HF Hartree Fock

HOMO highest occupied molecular orbital

IC internal conversion

i.e. that is

ISC intersystem crossing

KIE kinetic isotope effect

L lysine

LC long-range corrected

LCAO linear combination of atomic orbitals

LDA Local Density Approximation

LUMO lowest unoccupied molecular orbital

MAE mean average error

MC monte carlo

mCEC modified center of excess charge

MD molecular dynamics

MEP minimum-energy pathway

ML machine learning

MM molecular mechanics

MmCPDII class II photolyase from archaea *Methanosarcina mazei*

MO molecular orbital

MR multireference

MSE mean squared error

MTD metadynamics

NeoR neorhodopsin

NHE N-methylamide

NMR nuclear magnetic resonance

NN neural network

OM orthogonalization model

OM2/MRCI orthogonalization model 2/multi reference configuration interaction

OPC optimal point charge

PBC periodic boundary conditions

PCA principal component analysis

PCET proton-coupled electron transfer

PL/CRY photolyase/cryptochrome

PME particle mesh Ewald

PMF potential of mean force

POPC 1-palmitoyl-2-oleoylphosphatidylcholine

ppR proteorhodopsin

PSII photosystem II

PT proton transfer

QM quantum mechanical

QM/MM hybrid quantum mechanics/molecular mechanics

R arginine

RESP restrained electrostatic potential

RMSD root mean square deviation

RNR ribonucleic reductase

RSBH$^+$ protonated retinal Schiff base

SCC self-consistent charge

SCF self-consistent field

SRII sensory rhodopsin II

SIE self interaction error

SPC simple point-charge

SORCI spectroscopy oriented configuration interaction

SE Schrödinger equation

SMILES simplified molecular-input line-entry system

SOAP smooth overlap of atomic positions

STOs Slater-type Orbitals

TD time-dependent

TST transition state theory

UMAP uniform manifold approximation and projection

US umbrella sampling

vdW van der Waals

VR vibrational relaxation

W tryptophan

WHAM weighted histogram analysis method

wrt with respect to

wt wild type

WT-MTD well-tempered metadynamics

Y tyrosine, Tyr

ZDO Zero Differential Overlap

ZPE zero point energies

1 Introduction

The structure of a protein, together with its underlying sequence, determines its function and reactivity. This thesis illustrates how the local environment of the reaction center, such as nearby residues or solvent exposure, can tune quantum processes, focusing on two fundamental examples: proton-coupled electron transfer (PCET) and the modulation of light absorption.

PCET refers to reactions in which a proton transfer is directly coupled to an electron transfer. It occurs either through a concerted mechanism, characterized by simultaneous transfer, or in a sequential fashion, involving either proton transfer (PT) first, followed by electron transfer (ET), or vice versa. This coupling facilitates efficient charge movement between organo-metallic complexes and/or organic cofactors. PCET plays a central role in processes ranging from energy conversion (e.g., solar cells) to radical catalysis and enzymatic reactions. Prominent biological examples include photosystem II (PSII), prostaglandin H synthase, ribonucleic reductase (RNR), cytochrome P450, and galactose oxidase.[1] Experimental and theoretical studies[2–5] seek to unravel PCET mechanisms by analyzing thermodynamic and kinetic data, such as rate constants. The protein environment plays a critical role in this process, given its ability to either initiate the transfer or favor a specific directionality. The forward PCET reaction between tyrosines Y731/Y730 in subunit α of RNR is facilitated by a nearby glutamate residue, thereby illustrating the latter case.[6] These insights are key for understanding biological functions and identifying general principles that inspire new experiments and guide protein design for technical applications.

Retinal proteins are photoreceptors containing a retinal chromophore embedded in the protein matrix, enabling organisms to detect and respond to light. They are essential for vision, bioenergetics, and phototaxis, and are broadly categorized into two types: microbial (type I) and animal rhodopsin (type II). Channelrhodopsins (ChRs), light-gated ion channels belonging to type I rhodopsins, are particularly important in optogenetics. This technology allows precise modulation of neuronal activity by depolarizing or hyperpolarizing targeted cells using light. Therapeutic applications of ChRs include vision[7–9] and hearing[10–13] restoration, as well as the treatment of neural disorders.[14,15] Color tuning, which is the ability of retinal proteins to absorb light at different wavelengths, allows retinal proteins to cover a broad range of the visible spectrum. A molecular-level understanding of retinal photo-properties in a given protein environment, as well as its influence on color tuning, is essential for advancing optogenetic applications and designing improved photoreceptor proteins.

In biomolecular systems, not only the active molecule, such as retinal or the catalytic residues in PCET, is important, but also the surrounding protein environment. Consequently, the description of the interaction network between the reactive molecules (e.g., retinal, redox cofactors) and their environment at a molecular level provides valuable insights for understanding protein behavior. The protein matrix modulates reactivity through electrostatic interactions, hydrogen bonding, or solvation effects. Factors such as solvent exposure, nearby amino acids, and spatial arrangement of functional groups influence reaction mechanisms and free energy barriers in PCET.[16–18] For color tuning in retinal proteins, key factors include the planarity of the ring-chain system, electrostatic interactions with binding pocket residues, and the complex counterion-chromophore interplay.[19–21] Computational chemistry provides powerful tools to explore protein function and elucidate enzymatic mechanisms at the molecular scale. It allows for the determination of kinetic, thermodynamic, and spectroscopic data that complement and explain experimental findings.

However, computational studies of PCET in a biological context are challenging due to the need to accurately capture quantum effects and solvent influence,[17,22] while photoreceptor studies are complicated by large system sizes and the intricate electronic structure of their chromophores.[23] To address these challenges, multi-scale simulations are required. Molecular mechanics (MM) force-field methods describe the protein environment, while quantum mechanical (QM) calculations accurately model the electronic structure of the active site. Hybrid quantum mechanics/molecular mechanics (QM/MM) simulations bridge these scales, enabling the description of fundamental reactions in biomolecules. Even so, they remain computationally demanding, particularly for large and complex proteins such as RNR, DNA photolyases, and membrane-embedded retinal proteins.

The aim of this work is to develop and validate protocols that facilitate QM/MM simulations of the quantum processes outlined above. We propose a two-step strategy, starting with developing the methods first on toy and model systems, including biomimetic peptides and the retinal chromophore in various solvents. These simplified yet biologically relevant systems are well suited to validate the methodology. Once this step is complete, the workflows can be applied to full protein systems.

A major part of this thesis focuses on a protocol for describing proton and electron transfer in PCET, combining QM/MM with metadynamics simulations. In this approach, collective variables (CVs) are used to track both proton and electron transfer, enabling reconstruction of free energy surfaces (FESs). Sampling is enhanced by biasing the proton transfer CV, while a reweighting procedure allows evaluation of the electron transfer. These FESs provide mechanistic insight and allow estimation of reaction free energy barriers. The protocol was validated on minimal systems and biomimetic peptides, and subsequently applied to complex proteins, such as class II photolyase from archaea *Methanosarcina mazei* (MmCPDII) and animal-like cryptochrome from green alga *Chlamydomonas reinhardtii* (CraCRY). These studies also included cases involving three distinct partners, with separate acceptors for proton and electron transfer, and water-mediated PCET. Additionally, a variant of the protocol that also biases the electron transfer CV was tested for realistic QM/MM setups.

For color tuning studies, the retinal chromophore was simulated in various solvents and ion configurations to mimic protein binding pocket conditions. Spectroscopic properties such as excitation energies and oscillator strengths were calculated and used to train neural networks. These were then applied to retinal proteins like channelrhodopsin Chrimson, bacteriorhodopsin (bR), and bovine rhodopsin to predict absorption spectra and gain insight into the chromophore-protein electrostatic interplay. A detailed investigation of Chrimson's active site identified structural features contributing to its red-shifted absorption.

This thesis highlights the interplay between structural features and protein function and presents case-specific computational workflows designed to capture these relationships in complex biomolecular systems.

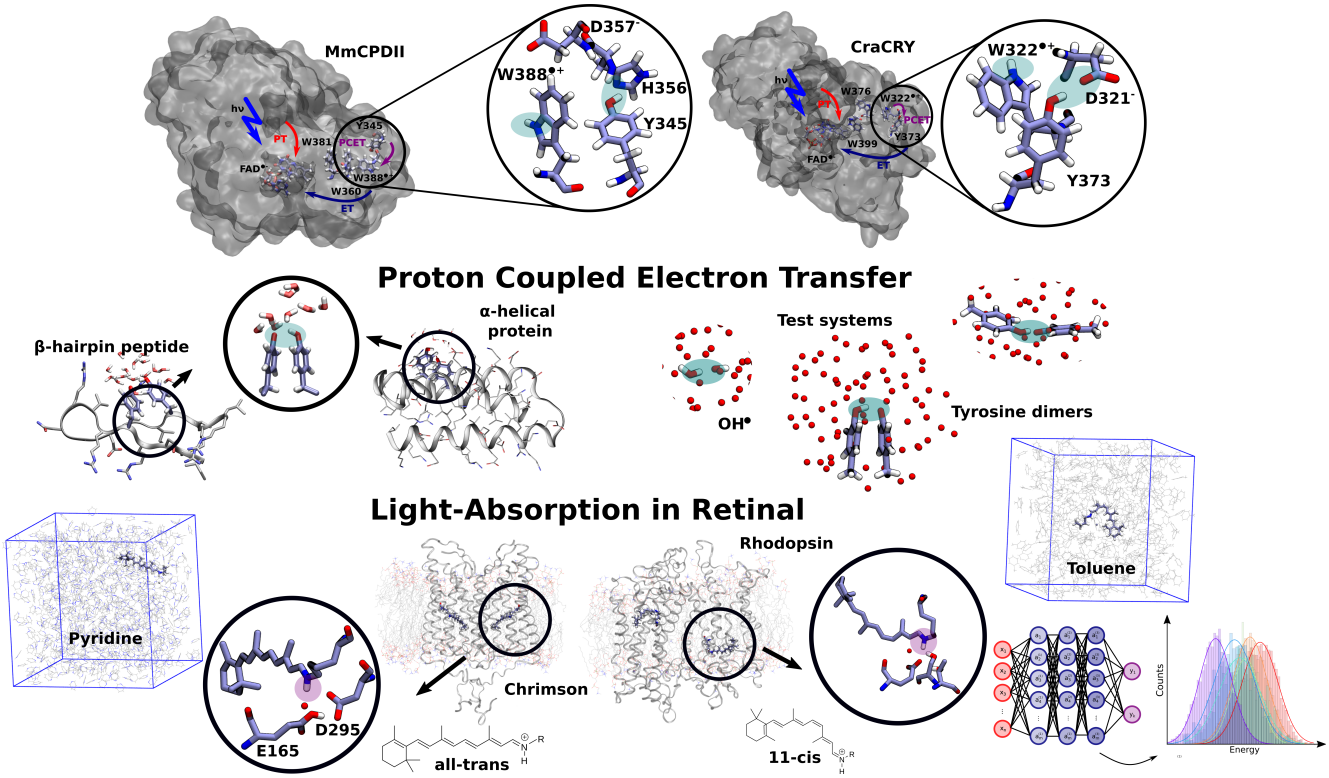


Figure 1.1: Overview of the molecular systems and topics investigated in this work. Upper part: Development of a protocol for describing PCET reactions, validated on test systems and biomimetic peptides, and subsequently applied to PCET reactions in two proteins, class II photolyase from archaea *Methanosarcina mazei* (MmCPDII) and animal-like cryptochrome from green alga *Chlamydomonas reinhardtii* (CraCRY). Lower part: Simulations of the retinal chromophore in various solvents and ion configurations to mimic the protein environment. Calculated spectroscopic properties were used to train a neural network, which was then applied to predict absorption spectra in different retinal proteins.

2 Theoretical Background

Computational chemistry provides various powerful tools to investigate molecules on an atomistic level. Fig. 2.1 gives a methodological overview. Quantum mechanical (QM) methods, based on *ab initio* calculations or density functional theory (DFT), can describe systems up to a few hundred atoms and are typically limited to femtosecond to picosecond timescales. Often, these systems are not propagated dynamically in a simulation, but instead optimized geometrically, followed by single-point calculations to evaluate electronic properties. Semi-empirical methods, such as density functional tight binding (DFTB), are based on QM principles but are accelerated through the use of empirical and tabulated values. These methods enable simulations of slightly larger systems and timescales up to several nanoseconds.[24]

For larger biomolecular systems with hundreds of amino acid residues and thousands of solvent molecules, molecular mechanics (MM) methods are necessary. These rely on empirical force fields and allow simulations to reach microsecond to millisecond timescales.[25] To achieve even longer simulations or to describe large assemblies, such as protein complexes or cellular components, coarse-graining models are employed. In this approach, several atoms are grouped into beads, reducing the number of particles and enabling the use of larger time steps.[26–28]

The larger the system, the greater the computational cost, which is why approximations are necessary, resulting in reduced accuracy. In the context of biomolecular studies, not only the reaction center is of interest, but also the environment, which modulates the reactions. Therefore, multi-scale approaches are crucial, as they combine the strengths of different computational methods. QM/MM simulations address the trade-off between accuracy and system size by treating the reactive region (such as the active site) with a QM or semi-empirical method, while the surrounding protein environment is described using a classical force field.[30,31]

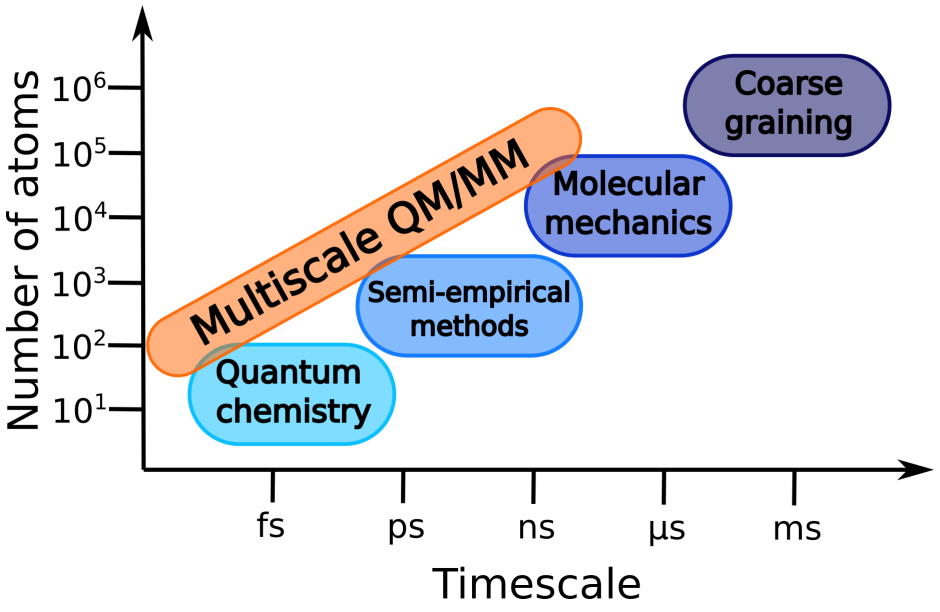


Figure 2.1: Methods in computational chemistry, adapted from Ref. [29]. Accuracy of the methods declines from left to right due to the use of more approximations, but larger systems can be considered. Hybrid hybrid quantum mechanics/molecular mechanics (QM/MM) simulations offer a balance between accuracy and efficiency by applying QM or semi-empirical methods to selected regions, while the remaining environment is described with MM.

2.1 Molecular Mechanics and Molecular Dynamics

2.1.1 Force Field Parameters

Eq. (2.1) describes the potential energy of a system according to MM force fields, which parametrize both bonded and non-bonded interactions. The first term represents the harmonic approximation of bond stretching E_{str}, followed by terms accounting for angle bending E_{bend} and torsional rotation around covalent bonds E_{tors}. The non-bonded interactions include Coulomb electrostatics E_{el} and van der Waals (vdW) interactions E_{vdW}. Coulomb interactions are modeled using effective atomic charges, while vdW interactions are described by the Lennard-Jones potential, which captures both Pauli repulsion and London dispersion

$$\begin{aligned} E_{\text{FF}} = &\underbrace{\frac{1}{2}\sum_{i}^{N_{\text{bonds}}} k_i(r_i - r_i^0)^2}_{E_{\text{str}}} + \underbrace{\frac{1}{2}\sum_{j}^{N_{\text{angles}}} k_j^{\vartheta}(\vartheta_j - \vartheta_j^0)^2}_{E_{\text{bend}}} + \underbrace{\frac{1}{2}\sum_{n}^{N_{\text{dihedral}}} V_n cos[n\omega - \gamma_n]}_{E_{\text{tors}}} \\ &+ \sum_{i}^{N}\sum_{j>i}^{N} \left[\underbrace{4\varepsilon_{ij}\left(\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right)}_{E_{\text{vdW}}} + \underbrace{\frac{1}{4\pi\varepsilon_0}\frac{q_i q_j}{r_{ij}}}_{E_{\text{el}}} \right], \end{aligned} \tag{2.1}$$

with N as number of particles. In E_{str} and E_{bend}, the parameters k_i and k_j^{ϑ} are force constants obtained from experimental data or quantum chemical calculations. r_i and ϑ_j represent the actual bond length and bond angle, respectively, while r_i^0 and ϑ_j^0 are their corresponding equilibrium values. The torsional term E_{tors} includes the amplitude V_n, periodicity n, and phase shift γ_n of the dihedral rotation. The Lennard Jones potential E_{vdW} describes non-covalent vdW interactions, comprising a short-range repulsive term (r^{-12}) and a longer-range attractive term (r^{-6}). Here, ε_{ij} represents the depth of the potential well, and σ_{ij} the distance at which the potential reaches its minimum. The Coulomb potential E_{el} models electrostatic interactions between point charges q_i and q_j, separated by a distance r_{ij}. These interactions are scaled by the dielectric constant ε_0.[25]

2.1.2 Newton's Equations of Motion for MD Propagation

Molecular dynamics (MD) simulations require the use of equations of motion, which are based on Newton's second law and must be solved numerically for many-body systems

$$\frac{d^2\vec{r}_i}{dt^2} = -\frac{1}{m_i}\frac{\partial V_{pot}}{\partial \vec{r}_i} = \frac{1}{m_i}\vec{F}_i \quad \rightarrow \quad m_i \cdot \ddot{\vec{r}}_i = \vec{F}_i. \tag{2.2}$$

In this context, $\vec{F}_i$ is the force acting on particle i, defined as the negative gradient of the potential energy V_{pot}. The acceleration $\ddot{\vec{r}}_i$ corresponds to $a(t)$, and m_i denoted the particle's mass. To begin a simulation, initial positions $\vec{r}_0$ and velocities $\vec{v}_0$ are required. These velocities are typically assigned according to a Maxwell-Boltzmann distribution at a given temperature. At each time step Δt, the forces and accelerations at t_i are calculated, followed by the computation of new positions and velocities at time t_i+Δt. This process is repeated iteratively.

Special integration algorithms, such as the Velocity Verlet integrator, are commonly employed to numerically propagate the system:[32]

$$r(t+\Delta t) = r(t) + v(t)\cdot\Delta t + \frac{1}{2}a(t)\cdot\Delta t^2, \tag{2.3}$$

$$v(t+\Delta t) = v(t) + \frac{1}{2}(a(t) + a(t+\Delta t)\cdot\Delta t. \tag{2.4}$$

There also exists the Leap-Frog algorithm.[33] These methods rely on Taylor expansions (typically up to the second-order term) and involve virtual time steps in both forward and backward directions to improve numerical stability and accuracy.

The time step Δt is a critical parameter in MD simulations and is typically chosen based on the fastest motion within the system. To ensure numerical stability, it is recommended that Δt be no greater than one-tenth of the period of the fastest vibrational mode. Since bond vibrations involving hydrogen atoms have characteristic periods of approximately 10 fs, a time step of $\Delta t = 1$ fs is commonly used.[25]

2.1.3 Geometry Optimization

In computational chemistry, the molecular structure with the lowest energy is of primary interest. To determine this minimum energy structure, the force acting on all atoms in the system must be calculated (see Eq. (2.5)). This is done by taking the derivative of the potential energy wrt the atomic coordinates in all spatial directions (x_i, y_i, and z_i)

$$\vec{F}_i = -\nabla_{\vec{r}_i} V_{\text{pot}} = -\begin{Bmatrix} \frac{\partial E}{\partial x_i} \\ \frac{\partial E}{\partial y_i} \\ \frac{\partial E}{\partial z_i} \end{Bmatrix}. \tag{2.5}$$

The structure is then iteratively adjusted by following the negative gradient of the potential energy surface until the energy no longer changes significantly between steps. This optimization process is considered complete once the energy change per iteration falls below a defined threshold, indicating convergence.[25]

2.1.4 Classical Equilibration and Production

2.1.4.1 Thermostat

Temperature control in MD simulations is achieved using a thermostat, which adjusts particle velocities by applying a scaling factor. One common approach is the Berendsen thermostat, which introduces an external heat bath with a defined reference temperature T_{ref} that is coupled to the molecular system. This coupling results in a gradual adjustment of the system's temperature toward the reference value, creating an NVT ensemble (constant number of particles N, volume V, and temperature T). The temperature scaling is performed every n_{TC} steps using the factor λ

$$\lambda = \left[1 + \frac{n_{TC}\Delta t}{\tau_t}\left\{\frac{T_{ref}}{T(t - \frac{1}{2}\Delta t)} - 1\right\}\right]^{\frac{1}{2}}, \tag{2.6}$$

with parameter τ_t closely related to the time constant τ of the temperature coupling. While the Berendsen thermostat does not generate a true canonical ensemble, as it suppresses proper fluctuations in kinetic energy, alternatives such as the Nosé-Hoover thermostat or the velocity rescaling method do. These methods introduce either additional dynamical variables or stochastic terms to ensure maintenance of the correct kinetic energy distribution.[25]

2.1.4.2 Barostat

To additionally control the pressure in the molecular system, barostats such as the Parrinello-Rahman pressure coupling can be used. In this method, the simulation box volume is adjusted dynamically using a coupling matrix that determines the strength of the pressure control. This allows the system to reach and maintain constant pressure, resulting in an NPT ensemble (constant number of particles N, pressure p, and temperature T).[25]

2.1.5 Increasing Computational Efficiency

2.1.5.1 Periodic Boundary Conditions

In MD simulations, the system should be kept as small as possible to reduce computational cost. At the same time, the molecule should be simulated in a bulk-like solvent rather than near an artificial surface. To achieve this, periodic boundary conditions (PBC) are applied, replicating the simulation box in all spatial directions. This creates the impression of an infinite system and minimizes surface effects.[25]

2.1.5.2 Treatment of non-bonded Interactions

Evaluating the non-bonded interactions is one of the most computationally demanding steps due to the extremely high number of possible atom-pair interactions. However, many of these interactions occur over large distances, so their contributions to the total energy are comparatively small. To reduce computational cost, the vdW potential is often truncated at a certain distance, such as 10 Å. For this purpose, a neighbor list is used, which stores atom pairs within the cutoff distance based on the initial geometry. This list is updated at regular intervals, e.g., every 20 steps.

The treatment of the other non-bonded interactions is more challenging, because Coulomb interactions decay as $\frac{1}{r}$ and therefore persist over longer distances than vdW interactions, which scale as $\frac{1}{r^6}$. Charge-charge interactions are inherently long-range, especially between species carrying net charges (e.g., ions), but can also be significant between atoms with partial charges in polar molecules. If a cutoff is introduced, the abrupt truncation can lead to discontinuities in the derivative of the energy function. To avoid this, a smooth switching function is applied that gradually reduces the interaction energy to zero near the cutoff distance. However, for accurate evaluation of electrostatic contributions, especially in periodic systems, the particle mesh Ewald method is often used. In this technique, the periodic system is separated into near and far field components. Near-field interactions are computed directly, while the far-field contributions are efficiently evaluated in reciprocal space using Fourier transform methods.[25]

2.1.5.3 Solvent Models

Water is a crucial solvent in biomolecular simulations. Compared to the number of atoms in the molecule of interest, the number of water molecules is typically very high. As a result, a significant portion of the computational effort is spent calculating interactions between water molecules. To reduce this cost while maintaining sufficient accuracy, simple but effective water models are used. Rigid water models, such as TIP3P[34] and simple point-charge (SPC),[35] apply constraints on bond length and angles, keeping them fixed during the simulation. In TIP4P[34] and optimal point charge (OPC)[36] water models have an additional interaction point to improve the description of electrostatic distribution.[25]

2.1.5.4 Bond Constraints and Time Step Control

Constraints on bond length are not only applied in water models but can also be introduced for bonds involving hydrogen atoms in other molecules. In such cases, algorithms such as LINCS,[37] SHAKE,[38] or RATTLE[39] are commonly employed. This reduces computational cost, primarily because the integration time step Δt can be increased: the fastest vibrational motions (those involving hydrogen atoms) are removed, and longer simulations become feasible.[25]

2.2 Sampling in Molecular Dynamics Simulations

2.2.1 Statistical Mechanics

Experiments are typically performed on a macroscopic scale (approximately 10^{20} particles), whereas computational simulations model microscopic systems (approximately 10^3-10^6 particles). Statistical mechanics bridges this scale gap by connecting the microscopic properties of individual particles with the macroscopic observables measured in experiments.

The Boltzmann distribution (Eq. 2.7) describes the probability P_i of a system occupying a particular energy state E_i, where k_B is the Boltzmann constant and Q is the partition function:

$$P_i = \frac{e^{-\frac{E_i}{k_B T}}}{Q}, \tag{2.7}$$

$$Q = \sum_{i=\text{states}}^{\infty} e^{-\frac{E_i}{k_B T}} = \int e^{-\frac{E_i}{k_B T}} \mathrm{d}r\,\mathrm{d}p, \tag{2.8}$$

which is defined or a system of N particles. Q is obtained by summing over discrete energy states E_i or by integrating over the full phase space, which includes all positions (r) and momenta (p). The partition function contains the statistical properties of the system and allows the derivation of thermodynamic functions such as internal energy U, Helmholtz free energy A, enthalpy H, entropy S, and Gibbs free energy G. These thermodynamic quantities can then be used to calculate macroscopic observables such as pressure P or the heat capacity at constant volume c_v, provided all energy states of the system E_i are known.[25]

2.2.2 Ensemble Averaging

MD and monte carlo (MC) simulations are performed in the condensed phase, meaning that a large collection of interacting particles is considered. For such systems, it is not possible to determine all energy states explicitly, and the partition function Q cannot be explicitly constructed. However, by generating a representative sample of the system, it becomes possible to estimate derivatives or differences of Q. This requires the construction of an ensemble, which is a collection of configurations that adequately represent the phase space relevant for the property of interest. If the ensemble provides a good sampling of the system, ensemble averages can be computed to yield the expected value of a property, denoted as $\langle X_M \rangle$

$$\langle X_M \rangle = \frac{1}{M} \sum_{i=1}^{M} X(E_i) = \frac{1}{M} \sum_{i=1}^{M} X(r_i, p_i). \tag{2.9}$$

These configurations may be described in terms of their energy E_i or their positions r_i and momenta p_i. The ergodic hypothesis states that the time average of a single particle is equal to the average over an ensemble of particles:

$$\langle A \rangle_{\text{time}} = \langle A \rangle_{\text{ensemble}}, \tag{2.10}$$

$$\langle A \rangle_{\text{time}} = \frac{1}{t} \int_0^t A(x, p, t')\,\mathrm{d}t', \tag{2.11}$$

$$\langle A \rangle_{\text{ensemble}} = \int\int \rho(x, p) A(x, p)\,\mathrm{d}x\,\mathrm{d}p. \tag{2.12}$$

This means that the system will eventually explore all relevant regions of phase space regardless of its initial state.

Energetic properties such as the internal energy U depend on the derivative of the partition function Q, and therefore sampling of low-energy regions of the phase is often sufficient. In contrast, entropic quantities such as the Helmholtz free energy A, entropy S, and Gibbs free energy G depend directly on Q, which includes exponential weighting of all energy states. This means that even rarely visited high-energy states can contribute significantly. As a result, sampling of the entire phase space

becomes necessary, which is computationally challenging. Methods like MD or MC tend to sample only low-energy regions, often causing the system to remain trapped in a limited volume of phase space. Consequently, calculating absolute values of entropic quantities is difficult. However, differences in these quantities, such as free energy differences, can still be determined reliably, as shown exemplarily for G[25]

$$\Delta G = G_A - G_B = -k_B T \left(\ln Q_A - Q_B\right) = -k_B T \ln\left(\frac{Q_A}{Q_B}\right). \tag{2.13}$$

2.2.3 Collective Variables

Collective variables (CVs) are functions of atomic coordinates designed to describe the essential motions or reaction paths of a molecular system. They are also referred to as reaction coordinates, as they ideally trace the progress along a specific transformation or reaction. The choice of a CV can range from simple geometrical descriptors such as bond length, angles, or torsions, to more complex constructs like combinations of distances, inverse distances, center of mass (COM), or even normal modes obtained from principal component analysis (PCA). Selecting an appropriate CV is often challenging, as it requires prior knowledge or strong hypotheses about the relevant reaction pathway or conformational transition.

In practice, CVs, s, can be monitored during or after MD simulations to analyze sampling and compute free energy surfaces (FESs). From the resulting probability distribution $P_0(s)$, the corresponding free energy $G(s)$ can be computed as

$$G(s) = -\frac{1}{\beta} \ln P_0(s), \tag{2.14}$$

with $\beta = \frac{1}{k_B T}$. For that to be true, the simulation must be unbiased and the relevant phase space must be adequately sampled.[25]

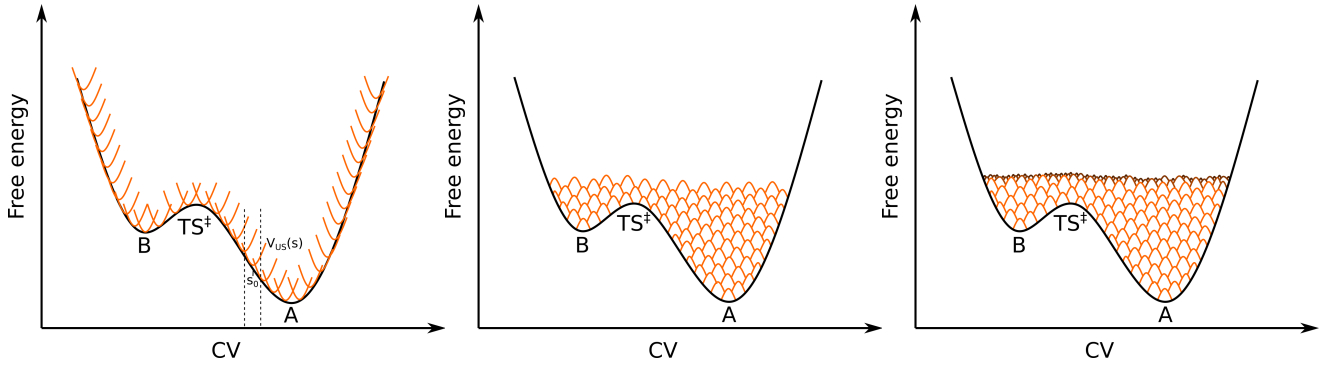
2.2.4 Enhanced Sampling

In many cases, more extensive sampling of phase space is required to obtain a reliable free energy profile along a reaction path as the same region of phase space is typically sampled multiple times, while new regions are explored only rarely. However, the timescale of the reaction of interest often exceeds that accessible by standard simulations. To overcome this mismatch, enhanced sampling methods are employed. In MD simulations, this can be achieved by parallel tempering approaches, where the region of interest is treated differently from the rest of the system, or by introducing a biasing potential that drives sampling toward specific regions of phase space. In the latter case, CVs are essential, as the biasing potential is applied directly to them; the following discussion therefore focuses on such biasing methods.

2.2.4.1 Umbrella Sampling

In umbrella sampling (US),[40] a so-called penalty approach is used, where a biasing potential $V_{\mathrm{US}}(s)$ in the form of a harmonic potential is centered at a chosen position s_0 along the reaction coordinate. Fig. 2.2(a) illustrates how the potential of mean force (PMF) is divided into overlapping intervals ("windows"), each explored by an individual simulation with the harmonic potential located at the window center.

To ensure continuous sampling across the reaction path, the force constant of the harmonic potential must be chosen such that adjacent windows overlap sufficiently. The complete PMF is then reconstructed by combining the biased distributions from all windows, typically using the weighted histogram analysis method (WHAM). However, the success of US depends critically on the choice of a suitable collective variable, since the convergence of the PMF is highly sensitive to the initial placement of the biasing potential.[25]



(a) Umbrella sampling **(b)** Metadynamics **(c)** Well-tempered metadynamics

Figure 2.2: Schematic representation of enhanced sampling methods, adapted from Ref. [41]. A: global minimum, B: local minimum, and TS‡: transition state. (a) Umbrella sampling: simulations are performed in overlapping windows, each restrained by a harmonic biasing potential. (b) Metadynamics: conformational space already visited is filled with repulsive Gaussian potentials, encouraging exploration of new regions. (c) Well-tempered metadynamics: the Gaussian height is gradually reduced to ensure convergence of the FES.

2.2.4.2 Metadynamics

Another enhanced sampling strategy is metadynamics (MTD),[42] which employs a time-dependent biasing potential (see Fig. 2.2(b)). In MTD, Gaussian-shaped biasing potentials are continuously deposited in regions of phase space already explored

$$V_{\mathrm{MTD}}(s,t) = \sum_{t_0=0,\Delta t,2\Delta t,\ldots}^{t-\Delta t} We^{-\frac{(s(t)-s(t_0))^2}{2\sigma^2}}, \tag{2.15}$$

with Gaussian height W and width σ as user-defined parameters, and Δt the fixed time intervals at which Gaussians are deposited. The height W determines the magnitude of the added bias, while the width σ controls the resolution of the sampling along the CV, often estimated from its fluctuations in an unbiased simulation. This progressively fills the free-energy wells, enabling the system to overcome energy barriers and sample new minima more efficiently. The accumulated biasing potential can then be used as an estimate for the underlying FES. Unlike US, MTD does not require detailed prior knowledge of the conformational landscape.

In standard MTD, the accumulated biasing potential V_{MTD} does not converge to a definite value, but instead fluctuates around the correct results. To overcome this limitation, well-tempered metadynamics (WT-MTD)[43] was introduced. In this approach, the initial height of the deposited Gaussians W_0 is rescaled and gradually reduced during the simulation

$$W = W_0\, e^{-\frac{V_{\mathrm{MTD}}(s,t)}{k_B \Delta T}}, \tag{2.16}$$

where $T+\Delta T$ is a fictitious CV temperature. The bias factor γ effectively controls how fast the Gaussian height decreases to zero

$$\gamma = \frac{T+\Delta T}{T}. \tag{2.17}$$

As a result, this approach not only resolves the convergence issue but also prevents overfilling of the FES (see Fig. 2.2(c)).[25]

Reweighting Procedure

Apart from the estimator presented above, the FES can also be calculated from the distribution of the CVs by applying a reweighting procedure. This approach recovers unbiased ensemble averages from a biased simulation, providing additional flexibility: distributions of CVs that were not explicitly biased during metadynamics can still be evaluated, and error estimates can be obtained. Reweighting further allows the extraction of thermodynamic quantities and the projection of the FES onto alternative CVs. It also facilitates convergence assessment, since the FESs obtained from summing the Gaussians and those reconstructed from reweighting should converge to the same result.

The biased probability distribution $P_{V_{\mathrm{MTD}}}(s(r),t)$ is related to the unbiased Boltzmann distribution $P(r)$ by:

$$P_{V_{\mathrm{MTD}}}(s(r),t) = P_{\mathrm{unbiased}}(r)\, e^{-\beta[V_{\mathrm{MTD}}(s(r),t)-c(t)]}, \tag{2.18}$$

$$c(t) = \frac{1}{\beta}\log\frac{\int \mathrm{d}s\, e^{-\beta G(s(r))}}{\int \mathrm{d}s\, e^{-\beta G(s(r))+V_{\mathrm{MTD}}(s(r),t)}}, \tag{2.19}$$

where $V_{\mathrm{MTD}}(s(r),t)$ is the time-dependent biasing potential applied during the simulation, and the time-dependent function $c(t)$ is often interpreted as an estimator for the reversible work introduced by the bias. Several algorithms exist to compute $c(t)$, either on-the-fly[44] during the MTD simulations or in a post-processing step.[45]

Using this relation, any observable $O(r)$ can be evaluated as an ensemble average over the unbiased distribution via reweighting

$$\langle O(r)\rangle_{\mathrm{unbiased}} = \left\langle O(r) e^{\beta[V_{\mathrm{MTD}}(s(r),t)-c(t)]}\right\rangle_{V_{\mathrm{MTD}}}. \tag{2.20}$$

In practice, the weights w used in the histogram analysis are computed as

$$w_{V_{\mathrm{MTD}}(s(r),t)-c(t)} = e^{\frac{V_{\mathrm{MTD}}(s(r),t)-c(t)}{k_B T}}. \tag{2.21}$$

The probability distribution from the weighted histogram analysis is then used to extract the FES, as described in Eq. (2.14).[44–46]

Error estimation

The methods above rely on the assumption that ensemble averages and histograms obtained from simulations provide a reliable estimate of the true population mean. However, MD simulations produce correlated data, which can lead to underestimated variances and, consequently, underestimated errors. This issue can be addressed through block averaging analysis: The trajectory is divided into multiple blocks, and the sample mean μ_i is computed independently for each. When the block size is too small, the resulting error estimate is too small, but it increases with block size and eventually plateaus and becomes statistically uncorrelated.[47] A useful approach for this type of analysis is the Bayesian Bootstrap method.[48]

2.3 Quantum Chemistry

Classical MM force fields, which rely on fixed bonding topologies, are insufficient to describe chemical reactions such as those in protein active sites. QM methods are required, as they can capture changes in the electronic structure - including bond breaking and formation as well as transitions between electronic states. There are also hybrid QM/MM methods, in which both realms are combined.

2.3.1 *Ab initio* Methods

QM methods are generally divided into two fundamental *ab initio* categories: wave function-based and electron-density-based approaches. Wave function methods explicitly describe the many-electron wave function, making them computationally demanding, with a cost that increases exponentially with the number of electrons. In contrast, electron density methods such as DFT are faster because the electron density of the ground state is described based on auxiliary one-electron Kohn-Sham orbitals. In comparison to the $3N$ coordinates of an N electron wave function, the ground-state electron density depends only on three spatial coordinates (plus spins).

However, the classification of *ab initio* methods depends on the definition. If restricted to wave function-based approaches without approximations beyond Hartree Fock (HF), DFT is excluded. This is mostly due to historical reasons, as the field of electron-density based methods emerged after the classification of *ab initio* and semi-empirical methods, which rely on empirical parametrization. If *ab initio* approaches are defined more broadly as any first-principles methods, DFT is included, which is the definition followed in this thesis.

2.3.1.1 Basic Principles of Electronic Structure Theory

The wave function Ψ can be obtained either from time-dependent, time-independent, or even the relativistic form of the Schrödinger equation (SE). The time evolution of Ψ can be obtained from the time-dependent SE, which takes the form of

$$\hat{H}(r,t)\Psi(r,t) = i\hbar\frac{\partial\Psi(r,t)}{\partial t}, \tag{2.22}$$

with i as imaginary unit, $\hbar$ as the reduced Planck's constant, and $\hat{H}$ as Hamilton operator. The square of the wave function Ψ

$$P(r,t) = |\Psi(r,t)|^2, \tag{2.23}$$

gives the probability $P(r,t)$ per unit of volume of finding the particle around position r at time t.

The time dependence can be separated from the spatial dependence of the wave function:

$$\Psi(r,t) = \psi(r)\exp\left(\frac{-iEt}{\hbar}\right), \tag{2.24}$$

with $\psi(r)$ as spatial wave function and E as energy in a phase factor. When considering time-independent problems, the latter term is neglected and the time-independent SE is used

$$\hat{H}\Psi(r) = E\Psi(r). \tag{2.25}$$

The Hamilton operator $\hat{H}(r,t)$ consists of the sum of the kinetic energy operator $\hat{T}$ and the potential energy operator $\hat{V}(r,t)$. Analytical solutions to this eigenvalue problem exist only for the simplest systems, such as the hydrogen atom. Therefore, approximate approaches are necessary once the system contains more than one electron.

The Born-Oppenheimer approximation simplifies the problem by decoupling nuclear and electronic motion

$$\Psi_{\text{total}}(\vec{r},\vec{R}) = \Psi_{\text{el}}(\vec{r},\vec{R})\phi_{\text{nuc}}(\vec{r},\vec{R}). \tag{2.26}$$

Nuclei are heavy in comparison to electrons and move more slowly, and can be treated as fixed. Therefore, electrons move in the field of static nuclei and the Hamiltonian simplifies: The nuclear kinetic energy operator $\hat{T}_N$ is neglected and nuclear repulsion $\hat{V}_{NN}$ is treated as a constant V_{NN} for a fixed nuclear geometry. The electronic Hamiltonian $\hat{H}_{el}$ for a system of N_e electrons and N_N nuclei results as

$$\hat{H}_{el} = \underbrace{-\frac{\hbar^2}{2m_e}\sum_i^{N_e}\nabla_i^2}_{\hat{T}_e} \underbrace{-\sum_A^{N_N}\sum_i^{N_e}\frac{e^2 Z_A}{r_{Ai}}}_{\hat{V}_{Ne}} + \underbrace{\sum_{i<j}^{N_e}\frac{e^2}{r_{ij}}}_{\hat{V}_{ee}} + V_{NN}, \tag{2.27}$$

with r_{Ai} denoting the distance between electron i and nucleus A, m_e is the electron mass, and ∇^2 represents the Laplace operator. Z_A and e, are the nuclear and elementary charges, respectively, and r_{ij} the distance between electrons i and j.

The energy of a trial wave function can be computed as the expectation value of the Hamiltonian operator, divided by the norm of the wave function

$$E_{\text{trial}} = \frac{\langle\Psi_{\text{trial}}|\hat{H}_e|\Psi_{\text{trial}}\rangle}{\langle\Psi_{\text{trial}}|\Psi_{\text{trial}}\rangle}. \tag{2.28}$$

The variation principle states that, for any normalized trial wave function Ψ_{trial}, this energy will always be greater than or equal to the exact ground-state energy E_0

$$\langle\Psi_{\text{trial}}|\hat{H}_e|\Psi_{\text{trial}}\rangle = E_{\text{trial}} \geq E_0 = \langle\Psi_0|\hat{H}_e|\Psi_0\rangle\,, \tag{2.29}$$

with Ψ_0 as wave function of the ground state. In practice, this optimization problem is solved by systematically varying the trial wave function to minimize E_{trial} and approaching E_0. This iterative procedure forms the basis of the self-consistent field (SCF) approach.[24,25,49,50]

2.3.1.2 Hartree Fock

The total electronic wave function Ψ is constructed as a combination of one-electron wave functions Φ, known as orbitals

$$\Psi(r_1, r_2, ..., r_N) = \Phi_1(1)\Phi_2(2)...\Phi_N(N). \tag{2.30}$$

To satisfy the Pauli exclusion principle, which states that no two electrons can occupy the same set of quantum numbers, the wave function must be antisymmetrical with respect to the interchange of any two electrons. This antisymmetry is achieved by representing Ψ as a Slater-determinant

$$\Psi_{SD} = \frac{1}{\sqrt{N!}}\begin{vmatrix} \Phi_1(1) & \Phi_2(1) & \cdots & \Phi_N(1) \\ \Phi_1(2) & \Phi_2(2) & \cdots & \Phi_N(2) \\ \vdots & \vdots & \ddots & \vdots \\ \Phi_1(N) & \Phi_2(N) & \cdots & \Phi_N(N) \end{vmatrix}, \tag{2.31}$$

in which the rows correspond to the electron coordinates and the columns to the occupied orbitals. HF is referred to as a mean-field approximation because the trial wave function is restricted to a single Slater determinant. As a result, electron correlation is neglected, and electron-electron repulsion is only accounted for in an average fashion.

The energy is minimized using the variation principle, which is used to derive the Hartree Fock (HF) equations with F as Fock operator acting on orbital Φ_i

$$F\Phi_i = \sum_j^{N_e} \lambda_{ij}\phi_j, \tag{2.32}$$

$$F = h_i + \sum_j^{N_e} (J_j - K_j). \tag{2.33}$$

Here, h_i is the kinetic energy of an electron and its attraction to all nuclei, while J is the Coulomb and K the exchange operator, both describing repulsive terms to all other electrons. λ_{ij} is a Lagrange function, which is stationary wrt orbital variations. When using canonical molecular orbitals (MOs) Φ'_i, Eq. (2.33) can be transformed into a set of eigenvalue equations

$$F_i\Phi'_i = \varepsilon_i\Phi'_i. \tag{2.34}$$

The MOs are expressed as linear combination of atomic orbitals (LCAO)

$$\Phi_i = \sum_{a=1}^{M_{basis}} c_{\alpha_i}\chi_\alpha. \tag{2.35}$$

When the HF equations are formulated in an atomic orbital basis, they take the form of the Roothan-Hall equations

$$FC = SC\varepsilon \quad \text{with } F_{\alpha\beta} = \langle\chi_\alpha|F|\chi_\beta\rangle \text{ and } S_{\alpha\beta} = \langle\chi_\alpha|\chi_\beta\rangle, \tag{2.36}$$

with Fock matrix F, the overlap matrix S, the coefficient matrix C, and the orbital energies ε_i. To solve the HF equations numerically, the SCF procedure is applied. A basis set of atomic functions χ_i and initial coefficients $c_{\alpha i}$ must be chosen to start the calculation and to form the Fock matrix F. Then F is diagonalized to obtain a new set of MO coefficients. This is iteratively repeated until convergence, which is when the set of coefficients used for constructing the Fock matrix is equal to those resulting from the diagonalization. Once the self-consistent wave function is determined, it can be used to compute various molecular properties.

The Hartree-Fock method neglects the electron correlation, so the HF energy E^{HF} always differs from the exact ground state energy E^{exact} by a certain correlation energy, defined as $E^{corr} = E^{exact} - E^{HF}$. To include electron correlation effects, a linear combination of Slater determinants is used in the configuration interaction (CI) method

$$\begin{aligned}\Psi_{CI} =& c_0\Psi_0^{HF} + \sum_{i=1} c_i\Psi_i \\ =& a_0\Phi_0^{HF} + \sum_S a_S\Phi_S + \sum_D a_D\Phi_D + \sum_T a_T\Phi_T + ... = \sum_{i=0} a_i\Phi_i.\end{aligned} \tag{2.37}$$

In contrast to HF, the CI expansion includes additional determinants representing single (S), double (D), triple (T), and higher excitations. These determinants are constructed by promoting electrons from occupied to virtual orbitals. To obtain the CI energy, the CI Hamiltonian matrix is constructed and diagonalized. The lowest eigenvalue corresponds to the CI ground state energy, and the associated eigenvector provides the coefficients a_i (see Eq. (2.37)). The second-lowest eigenvalue gives the energy of the first excited state, and higher eigenvalues correspond to further excited states.[25,50,51]

2.3.1.3 Density function-based Methods

Hohenberg and Kohn proved that the ground state electronic energy is a unique functional of the electron density ρ. This means that the entire system is fully characterized by its electron density. Specifically, the integral of the density determines the number of electrons, the cusps in the density reveal the positions of the nuclei, and the height of the cusps corresponds

to the nuclear charges. The first Hohenberg-Kohn theorem states that the external potential $V_{ext}(r)$, and thus the full many-body Hamiltonian, is uniquely determined by the ground-state electron density $\rho(r)$. As a result, all ground-state properties, including the ground-state wave function, are functionals of the electron density. Consequently, the total ground-state energy E is expressed as a functional of $\rho(r)$

$$E = E[\rho]. \tag{2.38}$$

The second Hohenberg-Kohn theorem establishes a variational principle: for any trial electron density $\tilde{\rho}(r)$ that corresponds to some valid N-electron system, the energy functional satisfies

$$E[\tilde{\rho}] \geq E_0, \tag{2.39}$$

where E_0 is the true ground-state energy. The energy functional $F[\rho]$ consists of three main contributions: the kinetic energy of the electrons $T[\rho]$, the electron-nucleus attraction $E_{ne}[\rho]$, and the electron-electron repulsion $E_{ee}[\rho]$. The latter can be further separated into a classical Coulomb term $J[\rho]$ and an exchange-correlation term $E_{xc}[\rho]$.

Modern DFT, as formulated by Kohn and Sham, introduces an auxiliary system of non-interacting electrons described by spin orbitals. This leads to the energy functional

$$E_{DFT}[\rho] = T_S[\rho] + E_{ne}[\rho] + J[\rho] + E_{xc}[\rho]. \tag{2.40}$$

This formulation improves upon the poor treatment of the kinetic energy in earlier orbital-free DFT methods. The kinetic energy is decomposed into two parts: the dominant contribution $T_S[\rho]$ and a smaller correction term absorbed into the exchange-correlation functional $E_{xc}[\rho]$. The first is that of non-interacting electrons in the Kohn-Sham potential. While numerically similar to HF kinetic energy of non-interacting electrons $\hat{T}_e$, they are conceptually distinct. The latter includes both exchange and correlation energies, as well as the correction to the kinetic energy

$$E_{xc}[\rho] = (T[\rho] - T_S[\rho]) + (E_{ee}[\rho] - J[\rho]). \tag{2.41}$$

The Kohn-Sham equation is solved iteratively using basis sets, analogous to the SCF procedure in HF theory. Unlike HF, which accounts only for exchange, DFT incorporates both exchange and correlation effects through the exchange-correlation functional $E_{xc}[\rho]$. This makes DFT more accurate and efficient for many systems, although there are cases where HF may outperform approximate DFT functionals, and vice versa. The expansion of MOs is based on integrating QM operators over a chosen set of basis functions. Generally, the larger and more flexible the basis set, the more accurate the resulting MO description. Common choices for basis functions include Slater-type Orbitals (STOs) and Gaussian-type Orbitals (GTOs), with the latter being favored in most practical applications due to computational efficiency.

DFT methods differ not only in the choice of basis sets but also in the formulation of the exchange-correlation functional E_{xc}. In the Local Density Approximation (LDA), E_{xc} depends solely on the local electron density, whereas in the Generalized Gradient Approximation (GGA) it depends on both the density and its gradient. Hybrid functionals, such as B3LYP, mix a fraction of exact HF exchange with DFT exchange and correlation. This balances accuracy across systems, due to error compensation: HF tends to overestimate certain properties, while DFT underestimates them.[25] The exchange-correlation potential V_{xc}, which is the derivative of the exchange-correlation energy E_{xc} wrt the electron density, should decay in a Coulombic fashion ($-\frac{1}{r}$) for $r \to \infty$. However, this asymptotic behavior is not correctly reproduced by LDA, GGA, or standard hybrid functionals. Long-range corrected functionals such as ωB97X address this limitation by partitioning the Coulomb interaction into short- and long-range components using the error function

$$\frac{1}{r_{12}} = \frac{1 - \mathrm{erf}(\omega r_{12})}{r_{12}} + \frac{\mathrm{erf}(\omega r_{12})}{r_{12}}. \tag{2.42}$$

In this approach, short-range exchange and correlation are described with DFT, where error compensation is effective, while the long-range contribution is treated with HF exchange, which partially mitigates the self-interaction error.[52]

Population analysis
Mulliken population analysis provides an estimate of partial atomic charges, i.e., the effective number of electrons associated with an atom A

$$q_A = \sum_i^{\mathrm{MO}} n_i \sum_{\mu \in A} \sum_B^{\mathrm{atom}} \sum_{\nu \in B} c_{\mu i} c_{\nu i} S_{\mu \nu}. \tag{2.43}$$

The net atomic charge Δq_A is then defined as the difference between the nuclear charge q_A^0 and the number of electrons q_A

$$\Delta q_A = q_A^0 - q_A. \tag{2.44}$$

In the case of Hirshfeld charges, atomic reference densities are used to construct the promolecular density $\rho_{\mathrm{promolecule}}(r)$, which is then employed to partition the molecular electron density at each point in space

$$\rho_{\mathrm{promolecule}}(r) = \sum_A^{M_{\mathrm{atoms}}} \rho_A^{\mathrm{atomic\,density}}(r). \tag{2.45}$$

The Charge Model 5 (CM5) extends this approach by applying empirical corrections to the Hirshfeld charges in order to better reproduce molecular dipole moments.

Another widely used approach is population analysis based on the electrostatic potential (ESP). Here, the ESP at position r is given by contributions from the nuclei and the electron density

$$\phi_{\mathrm{ESP}}(r) = \sum_A^{\mathrm{nuclei}} \frac{Z_A}{|r - R_A|} - \int \frac{\rho(r')}{|r - r'|} \mathrm{d}r' = \sum_A^{\mathrm{nuclei}} \frac{Q_A(R_A)}{|r - R_A|}, \tag{2.46}$$

the latter giving an approximation by assigning an atomic charge Q_A to atom A. Merz-Kollman population analysis is based on this principle.[25]

2.3.2 Semi-empirical Methods

The *ab initio* methods described above are typically applicable only to small molecular systems, often not exceeding on the order of ∼100 atoms, due to their steep computational scaling. Although these methods can be highly accurate, their computational cost is considerable. To address this, faster but still accurate approaches, known as semi-empirical quantum mechanical methods, were developed. These methods solve the SE using a combination of theoretical approximations and parameters derived from experiments or high-level *ab initio* calculations.

Semi-empirical methods approximate the electronic SE within a simplified HF framework, supplemented with empirical parameters. Efficiency is gained by (i) considering only valence electrons, (ii) employing a minimal basis set, and (iii) using the Zero Differential Overlap (ZDO) approximation. The ZDO approximation neglects products of basis functions centered on different atoms when they depend on the same electron coordinates, thereby eliminating multi-center electron repulsion terms. As a result, only one- and two-center integrals remain, which can be efficiently evaluated, enabling the practical use of STOs as basis functions.

The remaining integrals are either calculated directly from atomic orbitals, assigned values based on atomic properties such as ionization potential, electron affinities, or excitation energies; or they are fitted to a large set of experimental data. While this drastically reduces computational cost, it limits the applicability of the method to systems for which suitable parameters exist.[25]

2.3.2.1 Orthogonalization Models

The orthogonalization models (OMs) retain the ZDO approximation but account for the fact that orbital overlap can influence qualitative results. Instead of parameterizing the Fock matrix F directly, a symmetrically orthogonalized form of F' is used. They incorporate effects such as Pauli exchange repulsion, penetration effects, and core-valence interaction via an effective atom-pair potential added to the core-core repulsion term.

In OM1, orthogonalization corrections are applied to the one-center one-electron term of the core Hamiltonian. OM2 extends these corrections to the two-center one-electron terms. OM3 simplifies OM2 by omitting less significant correction terms. The latter two methods include three-center contributions in the correction of resonance integrals, improving modeling conformational properties. All OMs are based on GTOs and represent the resonance integrals in an empirical representation.[25,53,54]

2.3.2.2 Density functional tight binding

Density functional tight binding (DFTB) is a semi-empirical method derived from DFT by expressing the Kohn-Sham total energy in a tight-binding form. In this framework, only valence electrons are treated explicitly, while core electrons are represented by effective potentials. DFTB employs: (i) a minimal atomic basis set (valence orbitals only), (ii) a two-center approximation (matrix elements are computed only between basis functions on at most two atoms, a ZDO-like simplification that neglects explicit three-center and multi-center integrals), and (iii) pre-tabulated integrals fitted to DFT calculations. These approximations yield a computational speed-up of roughly three orders of magnitude compared to DFT.

The reference electron density ρ_0 is taken as a superposition of neutral atomic densities computed in advance. This eliminates the need for on-the-fly multi-electron integral evaluation. The total density is expressed as $\rho(r) = \rho_0 + \delta\rho(r)$, and the exchange correlation energy is expanded in a Taylor series.

In first-order DFTB (DFTB1), second and higher order terms in the Taylor expansion are neglected. With a minimal basis and two-center matrix elements, the total energy is[55,56]

$$E^{\text{DFTB1}} = \underbrace{\sum_{i}^{\text{MO}} n_i \sum_{a,b} \sum_{\mu\in a} \sum_{\nu\in b} c_{\mu i} c_{\nu i} H^0_{\mu\nu}}_{E^{\text{H0}}} + \underbrace{\frac{1}{2}\sum_{ab} V^{\text{rep}}_{ab}}_{E^{\text{rep}}}. \tag{2.47}$$

Second-order DFTB (SCC-DFTB or DFTB2) retains the quadratic term in the Taylor series, approximating it via atom-centered Mulliken charges

$$E^{\text{SCC-DFTB}} = E^{\text{H0}} + \underbrace{\frac{1}{2}\sum_{ab} \gamma_{ab} \Delta q_a \Delta q_b}_{E^{\gamma}} + E^{\text{rep}}. \tag{2.48}$$

The analytical function γ_{AB} describes the Coulombic interaction of atomic charge densities. It involves the atomic Hubbard parameters U_A that are related to chemical hardness. The charge-dependent Hamiltonian $H_{\mu\nu}$ is modified based on the variation principle during self-consistent charge (SCC) iterations to ensure convergence in the charge distribution.[57]

DFTB3 includes the cubic term in Taylor expansion, improving the description of polarization and hydrogen bonding

$$E^{\text{DFTB3}} = E^{\text{H0}} + E^{\gamma} + \underbrace{\frac{1}{3}\sum_{A,B} \Gamma_{AB} \Delta q_A^2 \Delta q_B}_{E^{\Gamma}} + E^{\text{rep}}. \tag{2.49}$$

The function Γ_{AB} includes the charge derivative of the Hubbard parameter U_A and γ_{AB}. The Hamiltonian is given as:

$$H_{\mu\nu} = H^0_{\mu\nu} + S_{\mu}\nu \cdot \Omega_{\mu\nu}, \tag{2.50}$$

$$\Omega_{\mu\nu} = \sum_C \Delta q_C \left(\frac{\gamma_{AC} + \gamma_{BC}}{2} + \frac{\Delta q_A \Gamma_{AC} + \Delta q_B \Gamma_{BC}}{3} + \frac{\Delta q_C (\Gamma_{CA} + \Gamma_{CB})}{6} \right), \tag{2.51}$$

with Ω_{AB} as atomic shift that is induced by charges. Refined parameter sets such as 3OB improve agreement with high-level calculations. A modified version, 3OB-f, was developed to yield better vibrational frequencies, though at the cost of reduced accuracy in describing energetics.[58–60]

Long-range corrected DFTB

Standard DFTB inherits short-range self-interaction errors from GGA exchange functionals. Long-range corrected (LC) DFTB mitigates these errors by splitting the electron-electron interaction into short- and long-range parts using an error function or exponential decay. This preserves error compensation in the short range while improving asymptotic behavior.[61–63]

Coupled-perturbed DFTB

The DFTB method has been extended with coupled-perturbed (CP) equations, which compute derivatives of atomic charges wrt atomic coordinates. This involves solving three sets of dependent quantities iteratively until SCC is reached:
(i) the derivatives of the MO coefficients,

$$\frac{\partial c^i_{\mu}}{\partial a} = \sum_m^{\mathrm{MO}} U^{(a)}_{mi} c_{\mu m}, \tag{2.52}$$

(ii) the derivatives of atomic charges,

$$\frac{\partial \Delta q_A}{\partial a} = \sum_i^{MO} n_i \sum_{\nu \in A} \sum_{\nu}^{\mathrm{AO}} \left(c_{\mu i} c_{\nu i} \frac{\partial S_{\mu\nu}}{\partial a} + \sum_m^{\mathrm{MO}} \left(U^{(a)}_{mi} (c_{\mu m} c_{\nu i} + c_{\nu i} c_{\nu m}) S_{\mu\nu} \right) \right), \tag{2.53}$$

and (iii) the derivatives of atomic shifts $\frac{\partial \Omega_{AB}}{\partial a}$.

The last term is not given in detail, but it takes all contributions from Eq. (2.51) into account. In a QM/MM setup, also the Hamiltonian shift induced by the ESP of the MM atoms must be taken into account:

$$\Omega^{\mathrm{QM/MM}}_{AB} = \Omega_{AB} + \frac{\phi_A + \phi_B}{2}. \tag{2.54}$$

This yields an additional set of CP-equations, involving the derivatives of $C_{\mu i}$, $\Omega^{\mathrm{QM/MM}}_{AB}$, and ΔQ_Awrt the coordinates of MM atoms b. These are solved in an iterative manner, similar to the approach above, facilitating accurate and consistent coupling within QM/MM simulations.[60,64,65]

This framework enables applications such as biasing atomic charges in metadynamics simulations:[66]

$$F_a = \frac{\partial V\left(S(\Delta q(\vec{r}))\right)}{\partial a} = \frac{\mathrm{d}V(S)}{\mathrm{d}S} \cdot \frac{\mathrm{d}S(\Delta q)}{\mathrm{d}\Delta q} \cdot \frac{\partial \Delta q(\vec{r})}{\partial a}. \tag{2.55}$$

2.4 Hybrid Quantum Mechanics/Molecular Dynamics Methods

As discussed earlier, an accurate description of chemical reactions such as bond breaking, bond formation, or electron transfer requires QM calculations. For large systems, such as biomolecules, a full QM treatment is computationally infeasible. Instead, hybrid quantum mechanics/molecular mechanics (QM/MM) methods combine a QM description of the chemically active region with an MM description of the surrounding environment (see Fig. 2.3). The QM region (e.g., the active site of a protein) is treated with either *ab initio* or semi-empirical methods, while the remainder of the system is modeled using a classical force field.

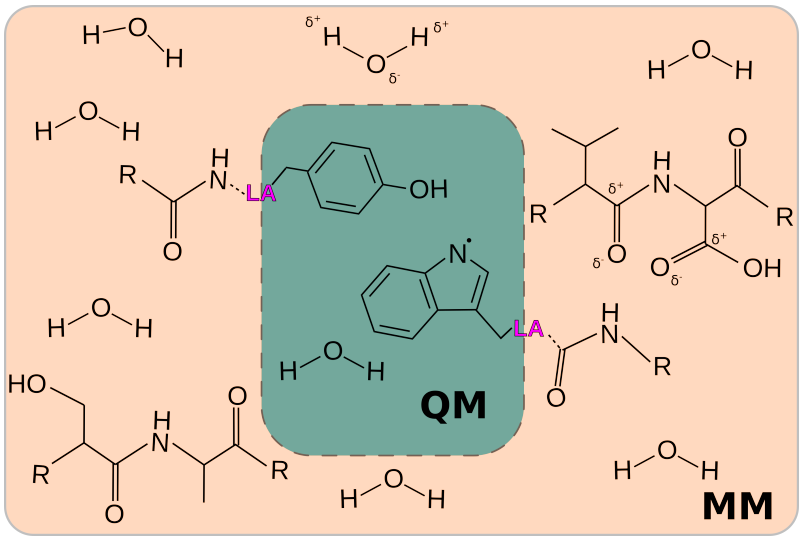


Figure 2.3: Schematic representation of the QM/MM concept (adapted from Ref. [31]). The active site is described using QM methods, while the surrounding environment is treated with an MM force field. The partial charges of the MM atoms interact with the QM region through Coulomb and vdW terms. The link atom (LA) concept is illustrated for capping covalent bonds that cross the QM/MM boundary.

The total energy in a QM/MM calculation can be partitioned into three contributions

$$E_{\text{total}} = E_{\text{QM}} + E_{\text{MM}} + E_{\text{QM/MM}}, \tag{2.56}$$

where E_{QM} is the QM energy of the active site or the reaction center, E_{MM} is the classical energy of the environment, and $E_{\text{QM/MM}}$ describes the interaction between the two regions. As in molecular mechanics, the QM/MM interactions are divided into bonded and non-bonded contributions (compare Eq. (2.1))

$$E_{\text{QM/MM}} = E^{\text{bonded}}_{\text{QM/MM}} + E^{\text{vdW}}_{\text{QM/MM}} + E^{\text{el}}_{\text{QM/MM}}. \tag{2.57}$$

The terms E_{MM}, $E^{\text{bonded}}_{\text{QM/MM}}$ and $E^{\text{vdW}}_{\text{QM/MM}}$ are obtained from the classical force field, while E_{QM} and the electrostatic coupling term $E^{\text{el}}_{\text{QM/MM}}$ must be evaluated explicitly.

The treatment of the electrostatic coupling $E^{\text{el}}_{\text{QM/MM}}$ depends on the chosen embedding scheme:

(i) Mechanical embedding: Electrostatic interactions between QM and MM regions are evaluated classically using fixed QM point charges Q_i and MM point charges q_j. The QM subsystem is computed in isolation (gas-phase Hamiltonian H_{gas}), so the MM environment does not polarize the QM electron density

$$E_{\text{QM}} + E^{\text{el}}_{\text{QM/MM}} = \langle \Psi_{\text{gas}} | H_{\text{gas}} | \Psi_{\text{gas}} \rangle + \sum_{i,j} \frac{Q_i q_j}{R_{ij}}, \tag{2.58}$$

where R_{ij} is the distance between QM atom i and MM atom j.

(ii) Electrostatic embedding: The MM point charges q_j are incorporated directly into the QM Hamiltonian H_{eff}, allowing the QM electron density to respond to the electrostatic environment

$$E_{\text{QM}} + E^{\text{el}}_{\text{QM/MM}} = \langle \Psi | H_{\text{eff}} | \Psi \rangle = \langle \Psi | H_{\text{gas}} | \Psi \rangle + \langle \Psi | \sum_{i,j} \left(\frac{Z_i q_j}{|R_i - R_j|} \int \frac{q_j \rho(r')}{|r' - R_j|} dr' \right) | \Psi \rangle \,, \tag{2.59}$$

in which Z_i and R_i are charge and position of the QM atom i, $\rho(r)$ the electron density of the QM system at point r, and R_j the position of the MM point charge q_j.

(iii) Polarizable embedding: MM atoms include static electric moments and polarizability; both regions polarize each other.

If the QM/MM boundary crosses a covalent bond, link atoms (often hydrogens) are used to cap the QM region to avoid unpaired electrons in the QM region. The charges of the truncated MM atom are set to zero, and its charge must be redistributed among neighboring MM atoms to avoid artificial overpolarization of the QM region. The selection of the QM region is critical: it must include all chemically relevant residues but remain small enough to keep calculations feasible.[25,31,67,68]

2.5 Machine Learning

Machine learning (ML) has rapidly transformed chemistry, enabling automated molecular design, property prediction,[29] and even breakthroughs in protein structure prediction, as exemplified by AlphaFold,[69,70] which was recognized with the 2024 Nobel Prize in Chemistry.[71] In chemistry, ML is now used to predict properties such as dipole moments, excitation energies, and reaction barriers. A particularly promising application is the development of machine-learned potentials, which aim to reproduce QM accuracy at a fraction of the computational cost. These ML potentials can be combined with MM in ML/MM approaches, offering an alternative to the QM/MM methods presented above.

An ML model is trained using a dataset with known input features (e.g., geometries or molecular descriptors) and a target property. Once trained, the model predicts this new property for new inputs, with accuracy assessed through error metrics and uncertainty estimates.

ML approaches in chemistry generally fall into two main categories:

- Supervised learning – a model is trained on labeled data, i.e., input features and known target values, to predict the target for new inputs. Each data point has a label, and the conditional probability can be expressed as

$$p(y|x) = \frac{p(x,y)}{\sum_{y'} p(x,y')}. \tag{2.60}$$

- Unsupervised learning – the aim is to identify patterns, clusters, or lower-dimensional relations in high-dimensional data, thereby revealing the most relevant features. In this case, there are no labels, only the distribution p(x), which can be factorized as

$$p(x) = \prod_{i=1}^{n} p(x_i|x_1,...,x_{i-1}). \tag{2.61}$$

These methods are increasingly important in chemistry, data science, and related fields. Their predictive power is remarkable, but reliability depends strongly on the quality and diversity of the training data. This makes dataset evaluation essential, especially for biomolecular systems.[29,72]

2.5.1 Introduction to ML Through Regression Models

ML models can be applied to various tasks, such as classification, regression, transcription, translation, outlier detection, or sampling. In chemistry, regression is particularly important for predicting continuous quantities such as reaction barriers. The goal is to construct a function

$$f(x_i) = b + \sum_{j=1}^{n} w_i x_{ij}, \tag{2.62}$$

where x_i are the input features, $f(x_i)$ the target values, w_i the weights, and b the bias. Here, the parameters w_i and b are optimized so that the predicted values $f(x_i)$ fit the training data $\hat{y}_i$ as closely as possible.

The dataset is typically split into training, validation, and test sets. The training process minimizes the mean squared error (MSE)

$$\mathrm{MSE}_{\mathrm{train}} = \frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - f(x_i))^2, \tag{2.63}$$

on the training set, but the ultimate goal is to minimize the MSE of the test set, known as the generalization error. Achieving a good balance between underfitting and overfitting is crucial. The capacity of a model is its ability to describe many different

functions. If it is low, underfitting occurs, which gives large errors on training data, while a high capacity leads to overfitting. This results in a large gap between the training and test error. Regularization introduces a penalty term to limit model capacity

$$J(w) = \mathrm{MSE}_{\mathrm{train}} + \lambda \sum_{j=1}^{n} w_j^2. \tag{2.64}$$

where λ controls the penalty strength. This reduces the generalization error while keeping the training error low. Optimization then also involves hyperparameters, such as λ and the learning rate ε, which are tuned by minimizing the validation error.

To report model performance, common statistical measures include the mean average error (MAE)

$$\mathrm{MAE} = \frac{1}{n} \sum_{j=1}^{n} |\hat{y}_i - f(x_i)|, \tag{2.65}$$

and the Pearson's correlation coefficient R^2,

$$R^2 = 1 - \frac{\sum_i^n (\hat{y}_i - f(x_i))^2}{\sum_i^n (\hat{y}_i - \bar{y})^2}. \tag{2.66}$$

where $f(x_i)$ is the predicted value and $\bar{y}$ is the mean of the reference value $\hat{y}_i$. Values of R^2 close to 1 indicate a strong agreement between predictions and reference.[72–74]

2.5.2 Descriptors

The molecular representation is crucial: it must provide all features necessary for learning while being invariant to translation, rotation, and permutation of identical atoms. Common representations include:[29,72]

- Simplified molecular-input line-entry system (SMILES) – a string-based notation convenient for storage and database indexing. While compact, it does not directly encode 3D structural information.
- Extended-connectivity fingerprints – graph representation that identifies atomic neighborhoods and bond types, mapped into fixed-length binary vectors.
- Coulomb matrix CM – encodes atom types and pairwise distances

$$CM_{ij} = \begin{cases} 0.5 Z_i^{2.4}, & i = j \\ \frac{Z_i Z_j}{|R_i - R_j|}, & i \neq j \end{cases} \tag{2.67}$$

  where Z_i is the nuclear charge, and R_i the position of atom i. Diagonal elements reflect element types, off-diagonals encode geometry. The CM is sensitive to atom indexing.
- Smooth overlap of atomic positions (SOAP) – represents the local environment of each atom as a smooth atomic density, expanded in a basis of Gaussian kernels.
- Graph Neural Network inputs – graph convolution of molecules occurs in message passing neural networks, where atoms and bonds are represented as vectors.

2.5.3 Neural Networks

Neural networks (NNs) are ML models composed of interconnected computational units called neurons a_i^j, which extract information from the input layers and produce complex output predictions (see Fig.2.4). In chemistry, the inputs x_n typically consist of molecular descriptors, while the output y_k corresponds to properties such as energy gaps or reaction barriers.

Each neuron combines input signals and passes them through a nonlinear activation function f. For example, consider neuron $a_2^{(2)}$ (orange in Fig. 2.4):

$$a_2^{(2)} = f\left(b_2^{(1)} + w_{2,1}a_1^{(1)} + w_{2,2}a_2^{(1)} + w_{2,3}a_3^{(1)} + w_{2,n}a_n^{(1)}\right) \tag{2.68}$$

$$= f\left(b_2^{(1)} + \sum_{i=1}^{n} w_{2,i}a_i^{(1)}\right), \tag{2.69}$$

where $b_2^{(1)}$ is the bias associated with this neuron, $w_{2,i}$ are the weights connecting layer 1 to this neuron, and $a_i^{(1)}$ are the activations of the previous layer. Common choices for f include the sigmoid, softplus, or the rectified linear unit. The weights w and bias b are the parameters learned from the training data. A loss function l (or cost function) measures the error between predictions $f(x_i)$ and reference targets $\hat{y}_i$. For instance, one typically uses the MSE

$$l = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{y}_i - f(x_i)\right)^2. \tag{2.70}$$

As in the case of regression, a regularization term to impede overfitting could be added to Eq. (2.70), similar to Eq. (2.64).

Error backpropagation efficiently propagates the prediction error through the network. By repeated application of the chain rule, the gradients of l wrt all weights w_{jk} and bias b_k are computed across all layers. Training is performed by minimizing l iteratively with a gradient descent algorithm in each epoch m

$$w_{ij}^{m+1} = w_{ij}^{m} - \varepsilon\frac{\partial l}{\partial w_{ij}} \tag{2.71}$$

where the learning rate ε controls the step size of the optimization process to update the network parameters and minimize the prediction error.[72–74]

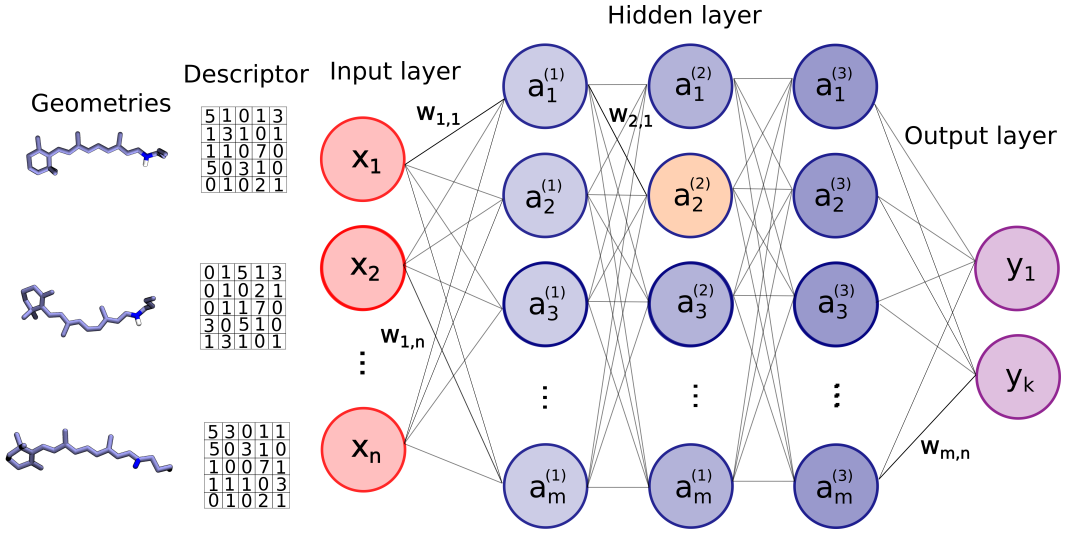


Figure 2.4: Schematic representation of a NN, adapted from Ref. [75]. Molecular geometries are first transformed into a descriptor, yielding feature vectors x_n that serve as inputs to the network. In the hidden layers, these features are processed through neurons where weights w_{jk} and bias b_k are optimized during training. The resulting transformed information is passed to the output layer, producing the predicted target values y_k.

2.6 Proton-coupled Electron Transfer

Proton-coupled electron transfer (PCET) is central to many energy conversion processes, such as photosynthesis and enzyme catalysis, where a proton transfer is coupled to an electron transfer in a redox reaction.

The coupling can proceed via two general pathways (see Fig. 2.5): (i) Sequential PCET, in which proton and electron transfer occur in separate, stepwise events, potentially passing through a thermally equilibrated and stable intermediate corresponding to a local minimum on the potential energy surface. (ii) Converted PCET, in which proton and electron transfer occur simultaneously, without the formation of such an intermediate.

Concerted PCET can be further classified into: (i) hydrogen atom transfer (HAT), where proton and electron are transferred together between the same donor and acceptor, leaving the overall charge distribution unchanged, and (ii) electron proton transfer (EPT), where the proton and electron move between different donors and/or acceptors, altering the charge distribution.

The solvent and surrounding environment can significantly influence PCET, leading to a variety of mechanistic scenarios that must be considered in theoretical descriptions. Marcus theory for electron transfer (ET) and related treatments for vibrationally nonadiabatic proton transfer form the basis of describing PCET.[76]

In this work, PCET is treated under the assumption of a vibronically adiabatic reaction, where the splitting between the ground electron-proton vibronic and the first excited vibronic state is much larger than the thermal energy $k_B T$. This allows the reaction to be described on the ground electronic surface with a classical treatment of the transferring proton.

Two CVs are used to describe PCET:

- Proton transfer CV – the antisymmetric stretch coordinate involving donor, acceptor, and proton

$$\Delta d = \mathrm{d}_{\mathrm{Donor-H}} - \mathrm{d}_{\mathrm{Acceptor-H}}. \tag{2.72}$$

- Electron transfer CV – the difference in total partial charge between donor and acceptor molecules, excluding the transferred proton

$$\Delta Q = \underbrace{\sum_{i}^{\text{mol \#1}} \Delta q_i}_{Q_{\mathrm{Donor}}} - \underbrace{\sum_{j}^{\text{mol \#2}} \Delta q_j}_{Q_{\mathrm{Acceptor}}}. \tag{2.73}$$

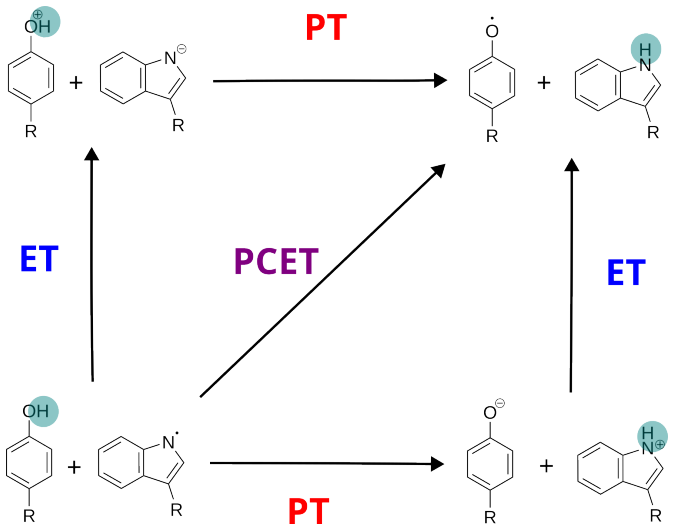


Figure 2.5: Scheme of PCET mechanisms exemplarily shown for a neutral tryptophan radical and a tyrosine residue. In concerted PCET, a simultaneous transfer occurs; whereas sequential PCET occurs either electron transfer first and the proton transfer (ET/PT) or the other way around (PT/ET).

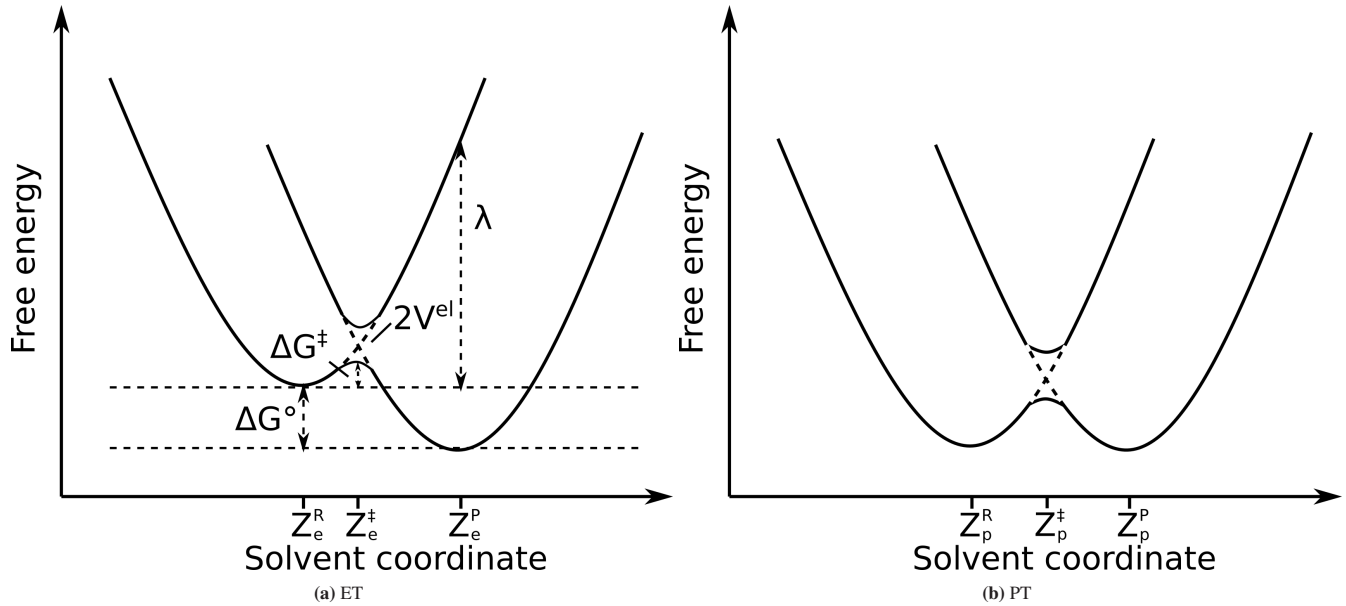
2.6.1 Electron Transfer

Electron transfer (ET) reactions can be described using a two-diabatic-state model, in which the reactant state corresponds to the electron localized on the donor, and the product state corresponds to the electron localized on the acceptor. The adiabatic electronic states are formed by diagonalizing the diabatic-state Hamiltonian along the reaction coordinate. These states can be represented as parabolic FESs along a collective solvent coordinate Z_e, which represents the combination of nuclear motions and solvent polarization that couple to the transfer (see Fig. 2.6(a)). The reaction is driven by motion along this coordinate, with the solvent configuration and polarization adjusting as the electron moves. This solvent reorganization, from the equilibrium geometry of the reactants to that of the products, is quantified by the reorganization energy λ. The driving force of the reaction, ΔG°, specifies whether the reaction is exergonic (spontaneous) or endergonic (uphill). In a symmetric reaction, the energy splitting between the adiabatic states at the crossing point is $2V^{\mathrm{el}}$, where V^{el} is the electronic coupling between the diabatic states. The reaction rate can be calculated using the Eyring equation

$$k_{ET} = \kappa \frac{k_B T}{h} \exp\left[\frac{-(\lambda + \Delta G^\circ)^2}{4\lambda RT}\right], \tag{2.74}$$

where κ is the transmission coefficient, accounting for dynamical effects such as recrossing of the transition state ($\kappa = 1$ in ideal adiabatic transfer, $\kappa < 1$ in nonadiabatic cases).

In electronically adiabatic reactions, the electronic coupling V^{el} is large compared to the thermal energy $k_B T$, allowing the electrons to adjust instantaneously to nuclear motion and solvent polarization. The reaction then proceeds entirely on the ground state adiabatic surface. In contrast, electronically nonadiabatic reactions occur when V^{el} is small in comparison to $k_B T$, and transitions between diabatic states involve higher electronic states. In this limit, Fermi's Golden rule can be applied to derive rate expressions for the nonadiabatic transition probability.[76–79]



(a) ET **(b)** PT

Figure 2.6: Schematic free energy curves as functions of the solvent coordinates, adapted from Ref. [78]. The crossing point of the parabolas is indicated with $Z^\ddagger$. Solid lines represent adiabatic free energy curves, and dashed lines represent the corresponding diabatic curves. (a) Single electron transfer: The reorganization energy λ is the energy required to reorganize nuclear positions and determines the curvature of the parabolas. ΔG° is the free energy difference between charge states. For a symmetric reaction, the intrinsic barrier $\Delta G^\ddagger \approx \lambda/4$. Twice the electronic coupling V^{el} corresponds to the energy gap between the ground and excited adiabatic surfaces ar the crossing. (b) Single proton transfer: Analogous free energy diagram for proton motion along the reaction coordinate.

2.6.2 Proton Transfer

Proton transfer (PT) plays a central role in many biochemical processes. In enzymatic acid-base catalysis, for example, proton motion underlies key reaction steps, while in bioenergetics protons are actively transported across membranes to generate a concentration gradient that drives adenosine triphosphate (ATP) synthesis. The rates of PT can be strongly modulated by the local environment near the reaction site.[80]

In most biochemical cases, PT is electronically adiabatic, meaning the electrons instantaneously adjust to the motion of the proton and the reaction proceeds on the electronically adiabatic ground state surface. Analogous to ET, the process can be modeled using two diabatic states and represented as parabolic free energy curves along a collective solvent coordinate Z_{P} (Fig. 2.6(b)), which incorporates the relevant nuclear motions of the proton, solute, and solvent.

If the proton is treated classically, the minimum-energy pathway (MEP) from reactants to products, passing through the transition state, defines the reaction mechanism and the barrier height. Rate constants can then be derived based on transition state theory (TST), with vibrational frequencies analysis enabling the inclusion of zero point energies (ZPE) and entropic corrections. In this classical treatment, nuclear quantum effects such as proton tunneling are neglected.

When the proton is treated quantum mechanically, classification schemes distinguish between vibrationally adiabatic and vibrationally nonadiabatic. In the adiabatic case, the proton remains in its vibrational ground state as the surrounding nuclei move, and large tunnel splitting can occur. In the nonadiabatic case, excited vibrational states of the proton participate in the transfer. The kinetic isotope effect (KIE), defined as the ratio of rate constants for hydrogen and deuterium transfer, provides an experimental signature of tunneling: large KIE values typically indicate significant hydrogen tunneling.[76,78]

2.6.2.1 mCEC Coordinate

The antisymmetric stretch coordinate, introduced in Eq. (2.72) as a CV for describing PT reactions, is well suited to direct PT. In some systems, however, long-range PT occurs via hydrogen-bonded water chains ("water-wires"), often with titratable amino acid side chains acting as donors or acceptors. Titratable amino acids are those whose side-chain protonation state can change depending on the pH of the environment. This includes aspartate (D), glutamate (E), histidine (H), arginine (R), lysine (L), and in some cases cystein (C) and tyrosine, Tyr (Y). These processes involve multiple bond-breaking and bond-forming events, as well as charge-separation, and require a quantum mechanical description of many atoms. In such cases, selecting an appropriate CV is essential to adequately sample the complex conformational landscape of the reaction. The modified center of excess charge (mCEC) coordinate ξ was developed for this purpose, representing the spatial location of the excess proton being transferred

$$\xi = \sum_{i=1}^{N_H} \mathbf{r}^{H_i} - \sum_{j=1}^{N_X} w^{X_j} \mathbf{r}^{X_j} - \sum_{i=1}^{N_H} \sum_{j=1}^{N_X} f_{\mathrm{SW}}\left(d^{H_i,X_j}\right)\left(\mathbf{r}^{H_i} - \mathbf{r}^{X_j}\right), \tag{2.75}$$

where N_H and N_X are the number of hydrogen and oxygen atoms, respectively, and $\mathbf{r}^{H_i}$ $\mathbf{r}^{X_j}$ denote their coordinates. The first term sums over all hydrogen coordinates. The second term is a weighted sum over the hydrogen-coordinating oxygen atoms, where w^{X_j} correspondents to the number of hydrogens coordinated to the oxygen X_j in the least protonated configuration. The third term provides a correction, using a switching function $f_{\mathrm{SW}}\left(d^{H_i,X_j}\right)$ applied to the distances d^{H_i,X_j} between hydrogens and their coordination oxygens:

$$f_{\mathrm{SW}}\left(d^{H_i,X_j}\right) = \frac{1}{1+\exp\left[\left(d^{H_i,X_j} - r_{\mathrm{SW}}/d_{SW}\right)\right]}, \tag{2.76}$$

where r_{SW} and d_{SW} are empirical parameters that control the center and steepness of the function. The vector ξ is projected onto a scalar CV ζ

$$\zeta = \frac{d_{\xi,D}}{d_{\xi,D} + d_{\xi,A}}, \tag{2.77}$$

taking the distances $d_{\xi,D}$ and $d_{\xi,A}$ between mCEC and either the initial donor or the final acceptor into account. By construction, ζ varies between 0 and 1: $\zeta = 0$ corresponds to the initial state with the excess proton located on the donor, while $\zeta = 1$ represents the product state with the excess proton located on the acceptor.[81,82]

2.6.3 Advanced Concepts in PCET

A more detailed description of PCET involves proton and electron nonadiabacity, which plays a central role in the theory of concerted PCET. The charge transfer reaction can be represented with two parabolic free energy curves along a collective solvent coordinate, which reflects the reorganization of the solvent and protein environment (see Fig. 2.7). Since not only the electron but also the transferring proton is treated quantum mechanically, the curves correspond to diabatic electron-proton vibronic states rather than purely electronic states, as in the Marcus theory described in section 2.6.1.

In this framework the mechanism proceeds as follows: first, the environment reorganizes toward the crossing point of the vibronic free-energy surfaces; at the crossing, a nonadiabatic transition occurs between reactant and product vibronic states (simultaneous tunneling of electron and proton); finally, additional environmental reorganization stabilizes the product.

Nonadiabaticity in PCET arises from three components: (i) the solute electrons, (ii) the transferring protons, both treated quantum mechanically, and (iii) the remaining nuclei, treated classically. This can be further classified into two cases: (a) vibronic nonadiabaticity, which concerns how electron and proton motions respond to the slower nuclei, and (b) electron-proton nonadiabaticity, which concerns how the electron responds to the proton motion.

Vibronic nonadiabaticity is characterized by the vibronic coupling V^{el}. As in the case of ET, when this coupling is much smaller than the thermal energy $k_B T$, Fermi's Golden rule can be applied to compute the rate constant. Electron-proton nonadiabaticity concerns the distinction between electronically adiabatic and nonadiabatic reactions. In electronically nonadiabatic reactions, the vibronic coupling is given by the product of the electronic coupling and the overlap integral between proton vibrational wavefunctions of reactant and product states. Nonadiabatic coupling manifests as a measurable change in the electron charge distribution such as change in dipole moment, ESP or partial atomic charges along the PT CV. This provides a way to discriminate between electronically adiabatic and nonadiabatic reactions, and thus to distinguish HAT and EPT. In the case of fast electron motion, as in HAT, the nonadiabatic coupling is small, since no significant redistribution of electronic charge occurs. However, when electron motion is slower, as in EPT, significant changes in the electronic charge distribution accompany the proton transfer.[83]

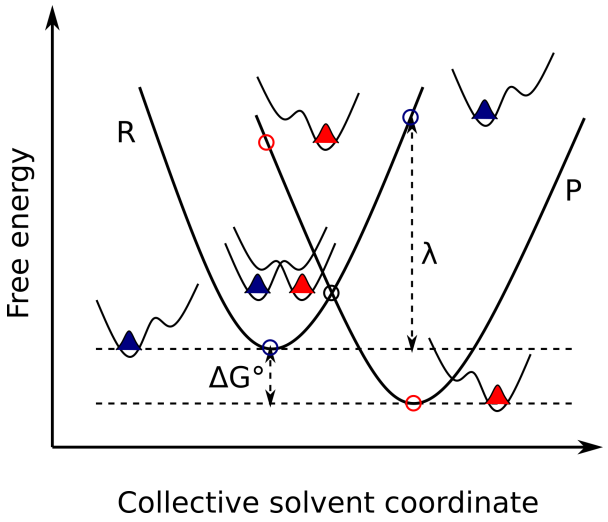


Figure 2.7: Schematic representation of parabolic electron-vibronic free energy curves along a collective solvent coordinate for a concerted PCET reaction in the electronic ground state, adapted from Ref.[83, 84]. Proton vibrational wavefunctions are shown in blue and red at both the reactant and product minima, with additional wavefunctions depicted at the crossing point.

2.7 Spectroscopic Properties

2.7.1 Light and Molecule Interaction

Photochemistry refers to chemical reactions initiated by the absorption of light. When a molecule absorbs radiation, the molecule is promoted from the electronic ground state S_0 to an electronic excited state S_n. The photon energy can then induce various physicochemical processes, such as photoisomerization, electron transfer, or fluorescence.

Light exhibits both wave-like and particle-like behavior. In the wave description, light in vacuum is represented as a linearly propagating electromagnetic wave consisting of orthogonal electric (**E**) and magnetic fields (**B**). In the quantum picture, light consists of photons, each carrying an energy

$$E_{ph} = h\nu = \frac{hc}{\lambda}, \tag{2.78}$$

where h is Planck's constant, ν the frequency, c the speed of light, and λ the wavelength. A photoexcitation occurs when E_{ph} matches the gap between two molecular energy levels. This process is commonly denoted as

$$A \xrightarrow{h\nu} A^*, \tag{2.79}$$

with A^* representing the electronically excited state S_n.

The probability of an electronic transition upon light absorption can be derived from the time-dependent SE (see Eq. (2.22)) within the Born-Oppenheimer (BO) approximation. In this framework, the total wave function Ψ_{total} can be separated into electron Ψ_{el} and nuclei ϕ_{nuc} contributions (see Eq. (2.26)). Interactions with light are mediated by the transition dipole moment μ, as the molecule interacts with an electric field

$$\hat{\mu} = -e\sum_i \vec{r}_i, \tag{2.80}$$

where e is the elementary charge and $\vec{r}_i$ connects the center of charge with the electron. The expectation value of the transition dipole moment of a transition between the ground state (0) and a final electronic-vibrational state (n) is given by

$$\begin{aligned}\langle\hat{\mu}_{0n}\rangle &= \langle\Psi^0_{\text{total}}|\hat{\mu}|\Psi^n_{\text{total}}\rangle &(2.81)\\ &= \langle\Psi^0_{\text{el}}|\hat{\mu}|\Psi^n_{\text{el}}\rangle \cdot \langle\phi^0_{\text{nuc}}|\phi^n_{\text{nuc}}\rangle &(2.82)\\ &= \varepsilon \cdot S. &(2.83)\end{aligned}$$

ε is the electronic transition dipole called the extinction coefficient and S is the Franck-Condon overlap integral between the lowest vibrational level in the ground state S_0 and any other vibrational level in an excited state S_n. The Franck-Condon factor S^2 depicts the state that overlaps best with the wave function in S_0 and is the strongest band in the absorption spectrum. The oscillator strength f, a dimensionless quantity describing the intensity of the transition, follows as

$$\begin{aligned}f_{0n} &= \frac{4\pi m_e}{3\hbar e^2}\nu_{0n}|\hat{\mu}_{0n}|^2 &(2.84)\\ &\propto |\mu_{0n}|^2, &(2.85)\end{aligned}$$

and is proportional to the square of the transition dipole moment.

The electromagnetic spectrum spans from high-energy γ-rays to low-energy radio waves. In the visible range, photon energies decrease from violet light ($\approx 420\,\text{nm} \equiv 2.95\,\text{eV}$) to red light ($\approx 700\,\text{nm} \equiv 1.77\,\text{eV}$).

Figure 2.8 shows the most common photophysical processes. A molecule in its singlet ground state S_0 can absorb a photon and be excited to a higher singlet state S_n. Due to the Franck-Condon principle, the transition occurs vertically on the

potential energy diagram because nuclear motion is negligible on the $\sim 10^{-15}$ s timescale of electronic excitation. The higher vibrational levels of S_1 or S_2 are often reached first due to greater wavefunction overlap.

From there, the molecule may undergo:

- Vibrational relaxation (VR): thermal relaxation to lower vibrational states (S_n to the lowest vibrational level of S_n).
- Internal conversion (IC): nonradiative relaxation to lower electronic states (S_2 to S_1).
- Fluorescence: radiative decay from the thermally equilibrated S_1 minimum to various vibrational levels of S_0, producing the vibronic structure in emission spectra.
- Intersystem crossing (ISC): spin-forbidden conversion to the triplet state T_1, followed by phosphorescence to S_0 on longer timescales.

What is not shown in the diagram, but also possible is:

- Photoisomerization ($A^* \rightarrow B$): A^* isomerizes to B and radiationless decay to the electronic ground state occurs. Therefore, the two potential energy surfaces come very close to each other, passing through the conical intersection, which is the avoided crossing region. As an example serves the isomerization in microbial rhodopsins from the all-trans to 13-cis isomer.
- Electron transfer ($A^* + B \rightarrow A^+ + B^-$): A^* is a better electron donor and acceptor than in the ground state, therefore an ion pair can be created. Photochemical PCET is an example of this primary photoprocess.

According to Kasha's rule, fluorescence originates from the lowest vibrational level of S_1, regardless of which S_n was initially populated. The fluorescence spectrum is often the mirror image of the $S_0 \rightarrow S_1$ absorption spectrum, but shifted to lower energy. The shift is called Stokes shift

$$\Delta E_{\text{Stokes}} = E_{\text{abs}} - E_{\text{em}} > 0. \tag{2.86}$$

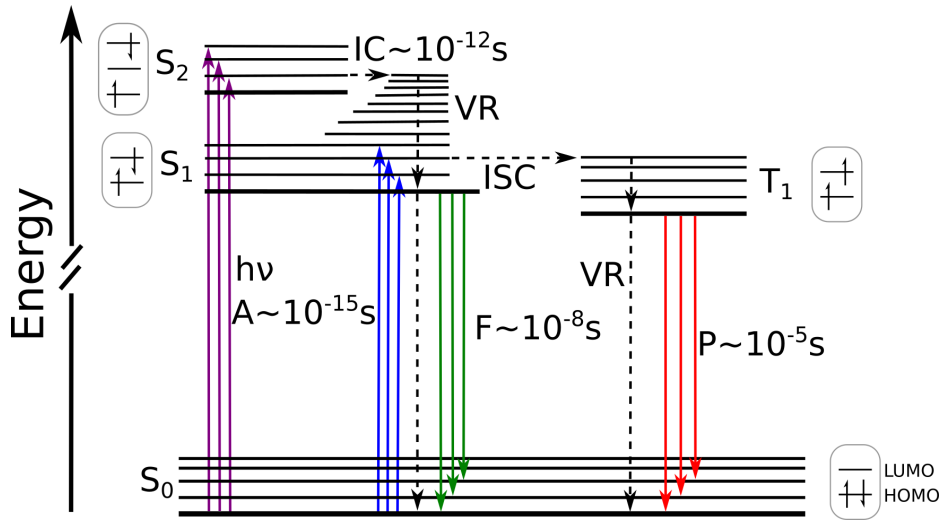


Figure 2.8: Jablonski diagram (adapted from Ref. [85]) illustrating key photophysical processes: The singlet ground S_0, first ,S_1, and second S_2 electronic states each contain multiple vibrational sublevels. Photon absorption A can promote the system from one orbital into a higher orbital, e.g. highest occupied molecular orbital (HOMO)–lowest unoccupied molecular orbital (LUMO) (S_1). Nonradiative vibrational relaxation (VR) leads to the lowest vibrational level of the excited state and internal conversion (IC) occurs to S_0, from which fluorescence emission F to S_0 or intersystem crossing (ISC) to the triplet state (T_1) may occur. This transition implies a spin inversion resulting in two unpaired electrons with the same spin. From T_1, phosphorescence to S_0 is possible, but slow due to the change in electronic spin. VR and IC typically occur on the picosecond scale, while fluorescence and phosphorescence are slower processes.

This shift results from vibrational relaxation in S_1 and emission into vibrational excited S_0 levels. Solvent effects, excited state reactions, and energy transfer can further increase ΔE_{Stokes}. VR is much faster than fluorescence, so the same emission spectrum is generally obtained, irrespective of the excitation wavelength.

It was stated above that a vertical transition from the ground state occurs in a higher vibrational level (Franck-Condon principle). This is indicated in Fig. 2.9. The reorganization energy λ is the energy difference between any higher lying vibrational state and the lowest vibrational level in the respective electronic state. The sum of the reorganization energies equals the Stokes-shift $\Delta E_{\text{Stokes}} = \Delta\lambda = \lambda^{S_0} + \lambda^{S_1}$.[85,86]

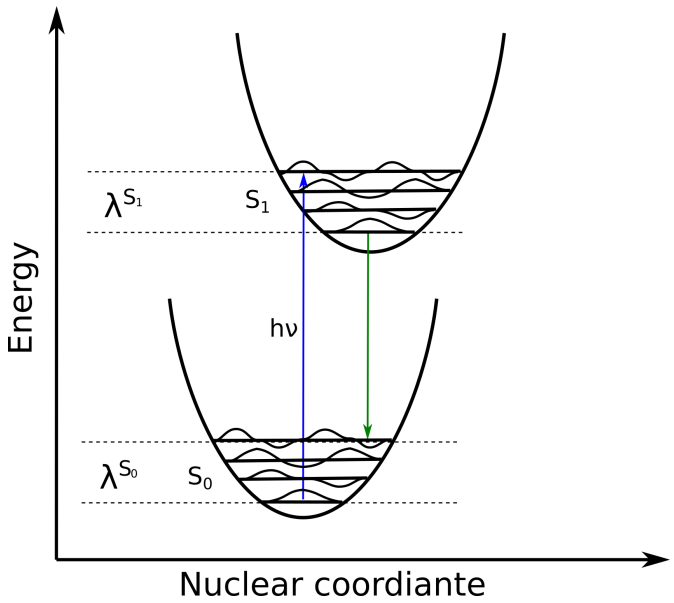


Figure 2.9: Schematic representation of the Franck-Condon principle, adapted from Ref. [86]: vertical transition (blue line) from S_0 to the highest vibrational level of S_1 due to greater wavefunction overlap. Vertical emission (green line) from the lowest vibrational level of S_1 to S_0. Reorganization energy λ is the energy difference between any higher lying vibrational state and the lowest vibrational level in the respective electronic state.

2.7.2 Treatment in QM Methods

If excited states are of interest, the aforementioned QM methods presented in section 2.3 must be adapted to describe also excited state properties.

In post HF-methods, one has to rely on multireference (MR) methods to explore both ground and excited state in a balanced manner. These methods are based on CI methods (see section 2.3.1.2), which write the molecular wave function as a linear combination of Slater determinants. Then electrons are taken out of occupied orbitals (so-called reference orbitals) into virtual orbitals to produce further Slater determinants. Then the final wave function includes not only single and double excitation, but also triple and quadruple excitations.[86,87]

2.7.2.1 Spectroscopy Oriented Configuration Interaction

Spectroscopy oriented configuration interaction (SORCI) is a MRCI based method, calculating energy differences between several electronic states in large molecules. Efficiency is gained by the use of a small reference space and SCF treatment, and other concepts combining variation and perturbation theory.[88]

2.7.2.2 Orthogonalization Model 2/Multi Reference Configuration Interaction

In semi-empirical frameworks, the OM methods presented in section 2.3.2.1 employ a MRCI treatment to account for static correlation effects in studies of excited states.[53,54,89]

2.7.2.3 Time-dependent DFT

Alternatively, excited state properties can be obtained using the time-dependent (TD) extension of DFT. In linear response TD-DFT, the system is described in its electronic ground state and subjected to a small, time-dependent perturbation, in this case an oscillating electric field. The perturbation induces changes in the electron density, from which excited state properties can be extracted.

In analogy to the Hohenberg-Kohn theorem, the Runge and Gross theorem establishes a correspondence between the time-dependent electron density $\rho(r,t)$ and the time-dependent external potential $V_{ext}(r,t)$, for a given initial state. This implies that the full time-dependent wavefunction $\Psi(r,t)$ is a functional of $\rho(r,t)$.

Linear-response TD-DFT starts from the ground-state Kohn-Sham orbitals and propagates the system to obtain the density response at a given perturbation frequency. The formalism requires an exchange-correlation functional $f_{\mathrm{xc}}(\mathbf{r},\mathbf{r}';\omega)$, which introduces an explicit dependence on the perturbation frequency ω.

The central working equation is the Casida equation, an eigenvalue problem of the form

$$\begin{pmatrix} A & B \\ B^{\dagger} & A^{\dagger} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \omega \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}, \tag{2.87}$$

where **A** and **B** contain ground-state KS orbital energies, orbital coefficients, and Coulomb and XC contributions. Solving this equation yields the excitation energies ω and the corresponding excitation vectors $(\mathbf{x},\mathbf{y})$, which can be used to compute oscillator strengths and transition dipole moments.[90–93]

2.7.2.4 TD-DFTB

In DFTB, an analogous linear-response approach is available for excited-state calculations. The single-particle energy differences ω are first obtained from the ground state DFTB calculation. A coupling matrix is then constructed (based on Mulliken transition charges and long-range Coulomb corrections). The eigenvalue problem is solved, yielding different spectroscopic properties, such as excited state dipole moments, non-adiabatic coupling vectors, symmetries, and the oscillator strength f of a transition.[94,95] A linear-response formulation is also implemented for long-range corrected exchange-correlation functionals, enabling a more accurate treatment of charge transfer reactions.[63]

2.7.3 Absorption Spectra Calculation

The methods described above provide access to spectroscopic properties by the means of theoretical methods. To compare these results with experimentally measured absorption spectra, it is important to account for the fact that a molecular absorption spectrum rarely consists of a single peak. Instead, it typically shows broad features, often with multiple peaks.

Only in very simple cases is it sufficient to use the vertical energy gap from S_0 to S_1 as the peak position of the experimental spectrum. In most cases, spectral broadening must be considered, arising from interactions with the environment and from transitions to vibrational levels other than the lowest one. Inhomogeneous broadening, e.g., due to structural fluctuations,

can be modeled by convolution with a Gaussian function, whereas homogeneous broadening, e.g., due to finite excited-state lifetimes, is more appropriately described by a Lorentzian function.

In the ensemble approach, molecular configurations are sampled from the ground-state potential energy surface, typically using molecular dynamics (MD) or monte carlo (MC) simulations. Spectroscopic properties are then computed for each snapshot, and the resulting set of excitation energies ω_i and oscillator strengths f is used to construct the spectrum

$$\sigma(\omega) = \sum_{i_{\text{frames}}} (f_i \cdot \omega_i). \tag{2.88}$$

The excitation energies ω_i are weighted by the oscillator strength f_i. The maximum peak is obtained by a Gaussian fit of the resulting histogram. This is valid because solvent-induced broadening is well represented by a Gaussian distribution. So, an environmentally induced homogeneous broadening in explicit solvent is considered in atomistic detail.[96]

3 Reweighting Metadynamics Simulations to describe PCET Reactions in Biomimetic Peptides

Chapter 3 is reproduced from Ref. [97] with permission from the PCCP Owner Societies:

- Spies, Katharina; Pfeffer, Leonie; Kubař, Tomáš; and Gillet, Natacha; Structural and environmental effects on the mechanism of biological proton-coupled electron transfer using DFTB/MM metadynamics., Phys. Chem. Chem. Phys. 27. 2025. pp. 15544–15556.

Author Contributions:
Leonie Pfeffer set up the structures for β-hairpin peptides H14 and W14 and prepared the systems using classical MD. Katharina Spies set up all other structures, performed hybrid quantum mechanics/molecular mechanics (QM/MM) molecular dynamics (MD) and metadynamics (MTD) simulations, and performed subsequent analysis. Katharina Spies, Tomáš Kubař, and Natacha Gillet contributed equally to the written text.

3.1 Introduction

Proton-coupled electron transfer (PCET) is a crucial redox reaction that occurs in a variety of processes, including energy conversion (solar cells), radical catalysis, and enzymatic reactions. Prominent PCET examples from biological systems include photosystem II (PSII), prostaglandin H synthase, ribonucleic reductase (RNR), cytochrome P450, and galactose oxidase.[1] They can involve a metal-center but also organic cofactors only, including residues such as, predominantly, tyrosine (Y), tryptophan (W), histidine (H) or glycine (G) and cysteine (C).

In its broadest definition, the term PCET describes the intertwined motion of a proton and an electron, regardless of the mechanism. It can occur simultaneously, i.e., in a concerted manner, or stepwise, i.e., sequentially, involving or lacking a reaction intermediate. Also, the donor and acceptor molecules can be different for the proton and the electron, so that three partners can be involved in the reaction.[76] The PCET mechanism strongly depends on the nature of the cofactors but can also be tuned by the close environment. For instance, long-range PCET in RNR is activated by the protein conformational change and its directionality over more than 30 Å is ensured by the close environment of the different tyrosines along the path.[17,98] Consequently, insights into the environment of the tyrosines and their dynamical behavior along the transfer process are important to understand PCET.

Both experimental and computational studies of complex proteins such as PSII and RNR are costly, time-consuming, and difficult to set up. An alternative way to control the environment of the active centers and explore the underlying PCET mechanisms turns out to be investigations on biomimetic peptides and model proteins. These facilitate, for example, the understanding of how non-covalent interactions affect the redox potentials of metal clusters or, as in our case, how the protein environment shifts the reduction potential and thus the electron transfer rate according to Marcus theory.[99,100]

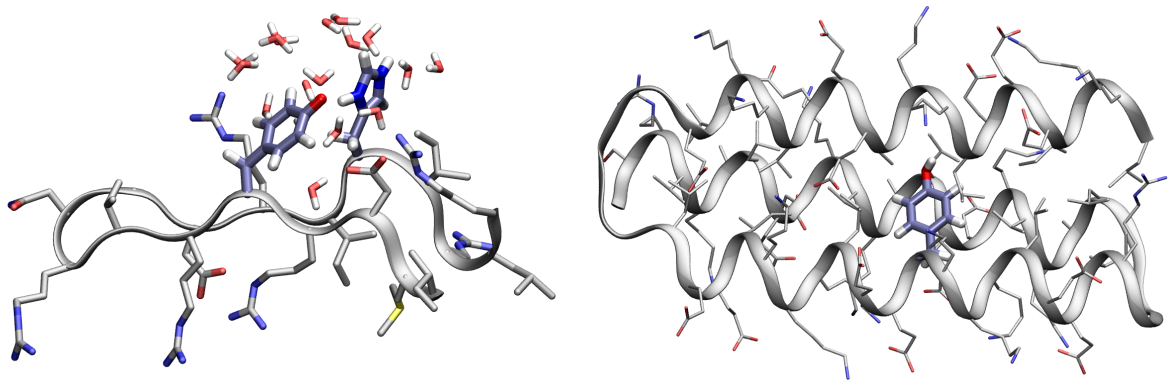
From that perspective, Sibert et al. designed an 18-residue β-hairpin structure in which an oxidized tyrosine can interact with a histidine residue via an interstrand PCET mechanism along a water chain (Fig. 3.1(a)).[99] Nuclear magnetic resonance (NMR) data confirm the stability of β-hairpin structure for the peptide and dipolar contacts between the histidine and tyrosine side

chains.[99] Electrochemical and electron paramagnetic resonance experiments show that the protein environment and the proton donors and acceptors control the midpoint potential and the reaction rate of the PCET. The oxidation of tyrosine leads to a PCET to histidine as the pK_a of histidine increases.[3,99,101] Experimental and computational analyses support a water-bridged PCET mechanism in this peptide. A series of mutant peptides was analyzed to investigate environmental factors that alter the mechanism.[3,101,102]

Besides, Tommos et al. synthesized a radical maquette α_3Y that folds into a stable α-helical structure (Fig. 3.1(b)).[103] Tyrosyl radicals that are generated in this maquette are incorporated into the protein structure and appear to be stable. The midpoint potential is increased at neutral pH compared to tyrosine in aqueous solution, indicating that the protein environment modulates the redox properties. The estimated lifetime for the tyrosyl radical is ≤ 30 ms, indicating that the protein environment strongly stabilizes the radical species. More recently, using a similar protein model, Nilsen-Moe et al. have described how the modulation of electron transfer (ET) and proton transfer (PT) driving forces, i.e., donor and acceptor redox properties and pK_a, modifies the PCET mechanism involving one tyrosine.[5] Indeed, tyrosine, thanks to its aromatic ring, has a relatively low ionization potential,[99,104] but its radical cation state presents a negative pK_a of –2.[105] However, the oxidized, deprotonated state of the tyrosine can be stabilized by the protein environment to lifetimes of up to 2.6 s.[106] Moreover, reactions following tyrosine oxidation employing fluorotyrosinated molecules have been studied computationally[6,107,108] and experimentally.[18,109] These studies demonstrate that the protein environment and its motions influence the reactivity and PCET mechanism.

Theoretical studies bring a deeper understanding of PCET mechanisms, by determining kinetic and thermodynamic properties with a system description at an atomistic level. QM/MM multiscale simulations are crucial to model PCET in biomolecules: the protein environment, which is described by molecular force field methods, can influence the donor and acceptor molecules that are involved in the transfer mechanism, which is described by quantum mechanics to represent the particle motions. PCET between aromatic amino-acid side chains was investigated computationally in several laboratories previously. Earlier studies were based on single-point calculations of the donor–acceptor systems. So, density functional theory (DFT) calculations (UB3LYP/6-31+G(d,p)) of a direct PCET reaction between the aromatic rings of Trp radical and Tyr with varying number of additional amino-acid residues separating the donor and acceptor in a peptide chain yielded activation energies of $E_a = 7-19\,\mathrm{kcal\,mol^{-1}}$.[110] Later studies obtained free energies by additionally including thermal corrections based on statistical mechanics; these were still single point calculations rather than averaging based on a sampling of configurational space however. A study of this kind estimated the activation free energy of the phenoxyl/phenol self-exchange reaction in a flipped conformation in the gas phase to be $5.6\,\mathrm{kcal\,mol^{-1}}$.[111] Another DFT study (D-B3LYP/def2-TZVPP) of a PCET between a pair of stacked tyrosine side chains in the gas phase yielded reaction free energy barriers of $\Delta G^{\ddagger} = 6.5$ and $11.0\,\mathrm{kcal\,mol^{-1}}$ for direct and water-mediated PCET processes, respectively.[112] It was also found that the effect of the electric field induced by the protein environment as point charges shifts the energies by only ca. $0.5\,\mathrm{kcal\,mol^{-1}}$. Most recently, PCET in biologically relevant complexes has been investigated by including sampling of configurational space with MD simulation. Such a study of PCET between two Tyr residues in RNR found $\Delta G^{\ddagger} = 4.2\,\mathrm{kcal\,mol^{-1}}$ for the side-chains in a stacked orientation and $9\,\mathrm{kcal\,mol^{-1}}$ in a flipped orientation, using QM/MM finite-temperature string simulations with umbrella sampling at a DFT level of theory (ωB97X-D/6-31+G**).[17] However, these methods are computationally expensive, so not always suitable for sampling the complex dynamical behavior of large biomolecules.

In this work, we propose a protocol based on QM/MM and MTD simulations in which we track collective variables (CVs) describing proton and electron transfer and recalculate the free energy surface (FES) to compare PCET in different biomimetic peptides using a methodology suitable for extended time scales and large biomolecules. Indeed, density functional tight binding (DFTB)3 and/or long-range corrected (LC)-DFTB calculations are 100 to 1000 times faster than those performed with full DFT, allowing longer simulations.[58,59] They can be combined with larger number of replicas in a multiple-walker MTD simulation setup for a qualitatively more thorough sampling of the conformational space of donors, acceptors and their environment. Here, a 1D MTD simulation is used, applying the extended sampling algorithm to the PT CV. The ET CV remains unbiased at runtime of the simulation and is instead only introduced in the post-processing of trajectories by reweighting.



(a) β-hairpin peptide H14 (wild type (wt))

(b) α-helical protein A (wt)

Figure 3.1: Structure of (a) β-hairpin peptide H14 (wt) and (b) α-helical protein A (wt) with the backbone shown as ribbons and the amino acid side chains displayed with emphasis on the residues involved in the PCET reaction and water molecules displayed in a 5 Å radius of the transferred hydrogen.

In fact, we have previously developed a methodology to apply bias to an ET CV in a 2D MTD simulation based on a DFTB version, in which coupled-perturbed (CP) equations are solved.[60,66] Although it would generate a FES of the CVs for PT and ET directly, that approach remains too computationally demanding so far.

We focus on the simulation of the two model systems shown in Fig. 3.1 to assess our computational protocol. We explore different donor-acceptor pairs and different configurations of the partners and the environment.

Apart from investigating the effects of the mentioned factors on the PCET process, this application has two rather technical aims: (i) It will be shown how a CV constructed as the difference of molecular charges can describe the charge state during the complex PCET process. (ii) It will provide reliable benchmark data that can be used later to validate the simulation protocol involving biasing potentials on the ET CV, as soon as the optimized CP-DFTB algorithms are finished.

3.2 Computational Details

3.2.1 Retrieving the FES from (un)biased MD Simulations

One or more CVs (s) must be introduced in order to describe the reaction of interest and to introduce the corresponding biasing potentials in metadynamics. In our case, the proton transfer (PT) is described by the difference of the distances of the hydrogen to the donor and acceptor (s_{PT}, Eq. (3.1)), where X is the heavy atom bound to the transferred proton (an oxygen or a nitrogen atom) and the electron transfer (ET) by the difference of the total charges (s_{ET}, Eq. (3.2)).[60,66]

$$s_{\mathrm{PT}}(\vec{r}) = \Delta d = d_{\mathrm{X_{Donor}H}} - d_{\mathrm{OX_{Acceptor}H}} \tag{3.1}$$

$$s_{\mathrm{ET}}(\Delta q) = \Delta Q = \underbrace{\sum_{i}^{\text{mol \#1}} \Delta q_i}_{Q_{\mathrm{Donor}}} - \underbrace{\sum_{j}^{\text{mol \#2}} \Delta q_j}_{Q_{\mathrm{Acceptor}}} \tag{3.2}$$

The Mulliken charges Δq_i are directly available in the deployed quantum mechanical (QM) method, DFTB, and the atomic charges are summed for each residue involved in the PCET. Each respective charge Q_{Donor} or Q_{Acceptor} includes the charges of side chain atoms up to and including Cβ and the hydrogens bonded to it (including the link hydrogen atom introduced in the QM/MM setup, see below), and it excludes any backbone atoms. The charge of the transferred proton is excluded from the difference. The CV for the proton transfer between two tyrosine residues consists of the difference between the proton-donor atom and the proton-acceptor atom, as shown in Fig. 3.2.

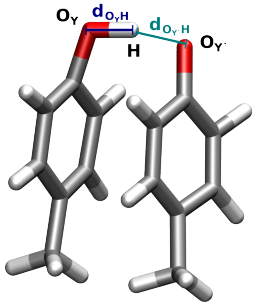


Figure 3.2: The proton transfer collective variables $s_{\mathrm{PT}} = \Delta d = d_{\mathrm{O_{Donor}H}} - d_{\mathrm{O_{Acceptor}H}}$ for the PCET reaction between two tyrosine residues.

We have performed two sets of simulations to represent the PCET FESs along these CVs: first, unbiased simulations where the proton is free to move; second, MTD simulations along the proton transfer coordinate. In the post-processing of the unbiased QM/MM MD simulations, these CVs are used to obtain a 2D histogram from the probability distribution $P_0(s) \equiv P_0(s_{\mathrm{PT}}, s_{\mathrm{ET}})$, which is also recalculated to the free energy $G(s)$ as

$$G(s) = -\frac{1}{\beta} \log P_0(s), \tag{3.3}$$

with $\beta = \dfrac{1}{k_B T}$ (T, thermodynamic temperature, k_B, Boltzmann constant).

The second set of simulations consisted of MTD that applied biasing potentials on the PT CV (s_{PT}). Standard MTD and well-tempered metadynamics (WT-MTD) were described in detail elsewhere.[46,113,114] An unbiased probability distribution has been retrieved from the biased one by means of reweighting.[44,45,115] A more complex reweighting procedure was required because (i) MTD uses a time-dependent biasing potential $V(s,t)$, and (ii) the desired histogram involved a CV that was not subject to the biasing potentials: those were applied to s_{PT}, and the goal was to estimate $P_0(s_{\mathrm{PT}}, s_{\mathrm{ET}})$.

Here, the reweighting factor[44,116]

$$c(t) = \frac{1}{\beta} \log \frac{\int \mathrm{d}s\, e^{-\beta G(s)}}{\int \mathrm{d}s\, e^{-\beta [G(s)+V(s,t)]}} \tag{3.4}$$

is an estimator for the reversible work done by the bias, and is evaluated by numerical integration as[44,46]

$$e^{\beta c(t)} = \int \mathrm{d}s \left[e^{\gamma V(s,t+\Delta t)/k_B \Delta T} - e^{\gamma V(s,t)/k_B \Delta T} \right] \tag{3.5}$$

where Δt is the time interval between the Gaussian depositions. This calculation was performed with PLUMED (version 2.5.1).[47,117] A reweighting factor based on $V(s,t) - c(t)$ makes it possible to estimate the equilibrium mean of any function by averaging along the MTD trajectory.

For instance, the unbiased distribution $P_0(s)$ can be obtained by reweighting the (biased) distribution $\hat{P}(s)$, obtained directly from the MTD simulation, with a factor of $e^{\beta[V(s,t)-c(t)]}$.

Importantly, it is possible to introduce an additional variable $s'(R)$ as a CV that, however, was not biased in the MTD simulation. This was used in our work to create 2D histograms $P_0(s_{\mathrm{PT}}, s_{\mathrm{ET}})$ by introducing an additional CV to describe the ET process: $s'(R) \equiv s_{\mathrm{ET}}(\Delta Q(R))$. The unbiased histogram (probability) was generated from the biased histogram (obtained directly from the metadynamics) using the quantity $c(t)$:

$$P_0(s,s') \propto \hat{P}(s,s') e^{\beta[V(s,t)-c(t)]}. \tag{3.6}$$

Practically, the `numpy.histogram2d` function was used to generate the distribution of both CVs, applying the weights $\beta(V - c(t))$ previously obtained from `rbias` by activating the `CALC_RCT` keyword in PLUMED. The resulting histogram was then converted to a FES using Eq. (3.3) and normalized by setting the minimum of ΔG to zero.

3.2.2 System Setup

The β-hairpin peptide structures are inspired by PSII and based on the amino acid sequence IMDRYRVRNGDRIHIRLR, first described by Sibert et al.[99] NMR data show that the wt peptide forms a well-ordered β-hairpin and that a PCET mechanism is possible between residues Y5 and H14. Since the NMR structures are not openly available, we based our structure on the solution NMR structure PDB ID 1KFP[118] of gomesin, an antimicrobial cysteine-rich peptide. We used the amino acid backbone coordinates from the 1KFP structure to reconstruct the β-hairpin structure of the target sequence using the leap program, which is part of the Amber program package.[119]

The α-helical protein structure inspired by the work of Tommos et al.[103] is based on PDB ID 2MI7.[109] The secondary structure schemes of the β-hairpin peptide and the α-helical protein are shown in Fig. 3.3, indicating the different mutations considered in this work. An overview of the biomimetic peptides investigated in this work is provided in Table 3.1 together with comments pointing out important differences between them.

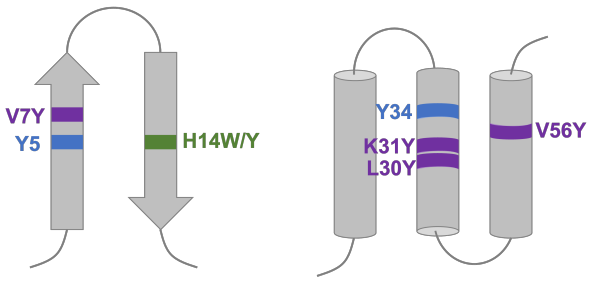


Figure 3.3: Secondary structure scheme of β-hairpin peptide (left) and α-helical protein (right) with emphasis on the residues involved in the PCET mechanism.

Table 3.1: Overview of the biomimetic peptides investigated in this work. $Tyr^{\bullet}$ is a deprotonated tyrosyl radical residue.

	$Tyr^{\bullet}$	PCET partner	comments
β-H14	Y5	H14 (wt)	$q_H = 1$
β-W14	Y5	H14W	
β-Y14s	Y5	H14Y	π-stacked conformation
β-Y14f	Y5	H14Y	flipped conformation
β-Y7	Y5	V7Y	
α-Y56	Y34	V56Y	embedded inside protein
α-Y31	Y34	K31Y	solvent exposed
α-Y30	Y34	L30Y	embedded inside protein

The wt and mutant structures of both systems are simulated using the Amber ff19SB force field and the SPC water model.[119,120] An additional set of simulations was set up, in which each of the two systems β-Y7 and α-Y30 was solvated using the OPC water model.[36] The atomic charges of the deprotonated radical tyrosine have been obtained using the restrained electrostatic potential (RESP) method as implemented in Antechamber[121,122] from the AmberTools software suite.[119] This new parametrization has been used to model the side chain Y5 (in the β-hairpin peptides) and that of Y34 (in the α-helical proteins) in a deprotonated radical state. An underlying electron density had been calculated on the HF/6-31G* level with Gaussian09.[123] The simulation box is cubic with a minimum solute to box distance of 12 Å and a salt concentration of 0.1 mM using an appropriate number of Na^+ and Cl^- atoms. Periodic boundary conditions were applied. The Verlet neighbor-searching scheme was used, the cutoff distance for the Lennard–Jones interactions was set to 12 Å, and the long-range electrostatic interactions were treated with particle mesh Ewald (PME).

3.2.3 Classical Equilibration and Simulation

The equilibration protocol consists of an energy minimization with a steepest descent algorithm until the maximum force dropped below $10\,000\,\mathrm{kJ\,mol^{-1}\,nm^{-1}}$ followed by a 60 ps NVT equilibration at 300 K using the SHAKE algorithm to constrain hydrogen–heavy atom bonding. The time step is 2 fs and Langevin dynamics is used to maintain the temperature of 300 K with a collision frequency of $5\,\mathrm{ps^{-1}}$. NPT equilibration was performed for 2 ns at 300 K using constant pressure and the SHAKE algorithm to constrain hydrogen bonding as before. The collision frequency is set to $1\,\mathrm{ps^{-1}}$. The production run is performed with the same parameters as the NPT equilibration for 200 ns with a time step of 2 fs.

The simulation protocol of the β-hairpins β-H14 (wt) and β-W14 (H14W) was slightly different. The steepest descent integrator was also used, performing a steepest descent minimization until the maximum force dropped below $1000\,\mathrm{kJ\,mol^{-1}\,nm^{-1}}$. NVT equilibrations were performed at 300 K for 100 ps, with a time step of 2 fs, using the leap-frog Verlet integrator. Bonds to hydrogen atoms were constrained using the LINCS algorithm. The NPT equilibration time was set to 10 ns with time steps of 2 fs. The reference pressure was 1.0 bar, and the Parrinello-Rahman barostat was introduced. All other simulation parameters were set as for the β-hairpin systems.

All of the classical simulations were performed using the GROMACS package (version 2021.5).[124,125]

3.2.4 QM/MM

Geometries with a maximum distance of 3 Å between the proton donor and acceptor atoms were selected from MM MD simulation to be used as starting structures for the QM/MM simulations.

All QM/MM optimizations and simulations were performed with the QM method DFTB3[58] as implemented in DFTB+,[65] using the standard 3OB parameter set.[59] The spin-polarized formalism of DFTB is used to describe the electronic structures with a radical character.[126]

The QM region consisted of the side chains of the two residues involved in PCET. To saturate the valence spheres of the relevant Cα atoms, hydrogen link atoms were introduced in the QM regions on the Cα–Cβ covalent bonds being cut by the QM–MM boundary. The MM region was described using the Amber ff19SB force field.[120] The distance between the proton donor and acceptor atoms was restrained with an upper wall restraint of 3 Å, and the sum of the distances between the hydrogen to each donor/acceptor atom was also restrained with an upper wall restraint of 3 Å using the PLUMED plugin (version 2.5.1).[47,117] The QM/MM simulations were performed for 1 ns with a time step of 0.5 fs. The QM/MM interface is the implementation in our local versions of DFTB+ and GROMACS.[127,128] The values of both CVs were extracted directly from the QM/MM simulation and used to compute the probability distributions and subsequently the FES according to Eq. (3.3). The ET CV was derived from Mulliken charges obtained using spin-polarized DFTB3/3OB.

3.2.5 Metadynamics

WT-MTD were performed within a multiple walker framework using a combination of GROMACS, DFTB3 as implemented in DFTB+, and PLUMED. The initial height of the Gaussians is set to $0.5\,\mathrm{kJ\,mol^{-1}}$ with a width $\sigma = 0.005$ Å and a bias factor of $\gamma = 6$. The frequency of the communication between the hill files was set to 500 steps, and the grid ranged from -4 to 4 Å. The keyword CALC_RCT was activated to obtain the reweighted bias and the factor $c(t)$[44]. The number of walkers varied from 8 walkers for the systems with two tyrosine residues to 48 walkers for the systems with tyrosine and histidine or tryptophan. The increase in the number of walkers was necessary to achieve convergence of the MTD simulations. All walkers were simulated with a time step of 0.5 fs for a total simulation time of 16 to 40 ns as required for the FESs to converge.

3.2.6 Recalculation of Electron Transfer Collective Variable

Additional single-point QM calculations were performed on selected snapshots from the unbiased QM/MM trajectories of β-Y14s and β-Y14f to assess the quality of the DFTB description of ET. The ET CV, ΔQ, was evaluated for the two Tyr side chains in vacuo, with the following methods, and subsequently compared: DFTB/3OB with and without the treatment of spin polarization, spin-polarized LC-DFTB[63] with the OB2 parameter set;[129] implementations in DFTB+ version 22.2 were used.[65] Full DFT on the levels M06-2X/6-311**[130–132] and ωB97X-D/6-311**;[133] these reference calculations were performed with Gaussian 09.[123] The ET CV is calculated from the atomic partial charges according to Eq. (3.2). In DFTB, the ET CV was computed using Mulliken atomic charges, and the CM5 correction was applied optionally. With the DFT methods, the ET CV was evaluated using either Mulliken or CM5-corrected Hirshfeld or Merz–Kollman (MK) population analysis.

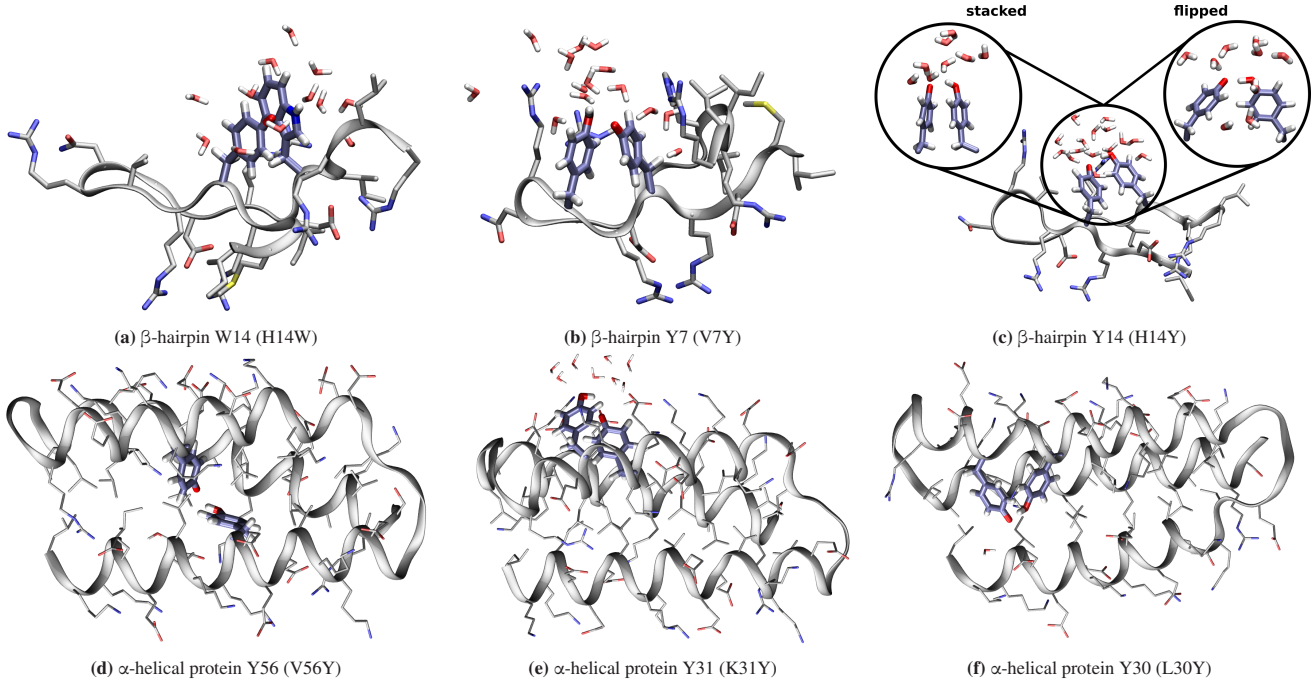
Based on this comparison, the Mulliken charges from single-point spin-polarized LC-DFTB/OB2 calculations, treating the protein environment as point charges, were selected as the representation of atomic charges to be used in Eq. (3.2). Thus, a total of 4000 snapshots per walker from the QM/MM MTD simulations were used for subsequent calculation of the ET CV ΔQ with this method. This is the first step in the post-processing of the MTD trajectories before the procedure according to Seq. 3.2.1 was carried out: build 2D histograms (using additionally the values of the PT CV Δd) and reweight to yield a 2D FES.

All data was processed and visualized with python[134] using the matplotlib,[135] NumPy,[136] pandas,[137] and seaborn[138] libraries.

3.3 Results

3.3.1 Secondary Structure of the Biomimetic Peptides

We study several biomimetic peptides (see Fig. 3.4), which differ in their secondary structure (β-hairpin peptides or α-helical proteins), the residues involved in the PCET mechanism (deprotonated radical tyrosine (Tyr•) paired with histidine (His), tryptophan (Trp), or tyrosine (Tyr)) as well as their relative spatial arrangement and solvation environment (either solvent exposed or buried within the protein).



(a) β-hairpin W14 (H14W) **(b)** β-hairpin Y7 (V7Y) **(c)** β-hairpin Y14 (H14Y)

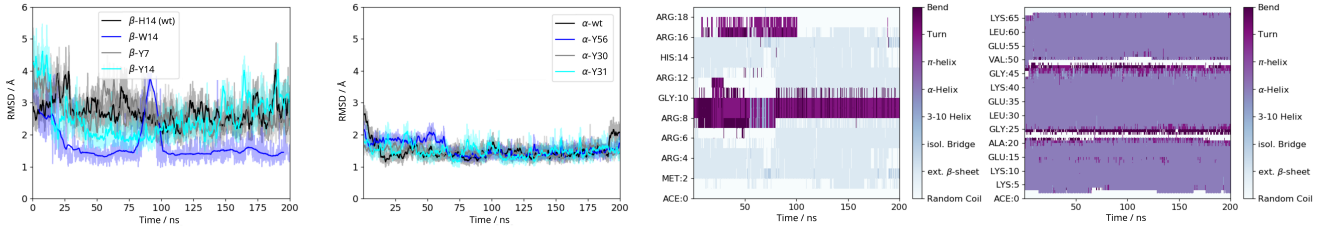
(d) α-helical protein Y56 (V56Y) **(e)** α-helical protein Y31 (K31Y) **(f)** α-helical protein Y30 (L30Y)

Figure 3.4: Structure of the β-hairpin peptides and α-helical proteins with the backbone shown as secondary structure and the amino acid side chains displayed with emphasis on the residues involved in the PCET reaction and water atoms displayed in a 5 Å radius of the transferred hydrogen.

In wild-type peptide β-H14 (WT) (Fig. 3.1(a)), the proton can transfer between Y5 and H14 either directly or via a water chain.[101] In peptide β-W14 (Fig. 3.4(a)), the residue H14 is replaced by tryptophan (W), as tyrosine/tryptophan dyads are conserved structural motifs in various proteins.[102,139] Examples are found in the β2 subunit of RNR, with a tyrosine and a tryptophan in a staggered T-shaped arrangement,[98] in PSII,[140] as well as in oxidoreductases,[141] cryptochromes[142] and photolyases.[143] Finally, peptide β-Y14 has H14 mutated to tyrosine in order to induce a symmetrical PCET reaction between the opposite β-strands. The V7Y mutation in peptide β-Y7 (Fig. 3.4(b)) is supposed to permit a PCET mechanism between two tyrosine residues on the same strand of the β-sheet. The insets in Fig. 3.4(c) show two different orientations of the tyrosines in the β-hairpin peptide β-Y14: a π-stacked (β-Y14s) and a flipped (β-Y14f) conformation. These geometries respectively correspond to the orientations of αY730/αY731 and αY731/βY356 found in RNR.[144]

The secondary structure of the α-helical protein A (wt) features three α-helices, as determined by CD and NMR spectroscopy.[103] There is only one tyrosine residue in the center of the wt sequence (Y34, Fig. 3.1(b)), which was thought to be buried in the protein structure.[18,109] By contrast, the protein structure changes during our simulations, exposing Y34 to the solvent, and the same event occurs in the α-Y31 (K31Y) variant (Figure 3.4(e)). The tyrosines in the α-Y56 (V56Y) and α-Y30 (L30Y) variants (Figures 3.4(d) and 3.4(f)), however, remain buried within the three α-helices structure.

The secondary structure conservation of the biomimetic peptides was evaluated by root mean square deviation (RMSD) analysis in combination with the Deffine Secondary Structure of Proteins (DSSP) algorithm. Fig. 3.5 shows the RMSD values relative to the average structure of all β-hairpin peptides (Fig. 3.5(a)) and α-helical proteins (Fig. 3.5(b)). All heavy



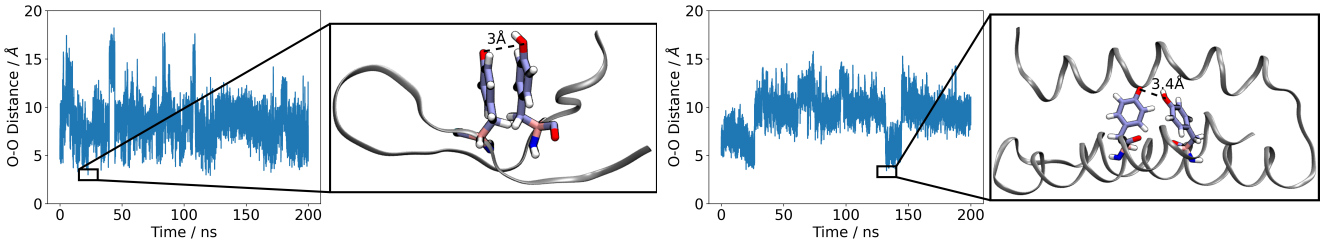
(a) H14 (wt), W14, Y7 and Y14 **(b)** A (wt), Y56, Y30 and Y31 **(c)** β-hairpin H14 **(d)** α-helical protein A

Figure 3.5: Secondary structure is conserved in all biomimetic peptides. RMSD Plots of (a) β-hairpin peptides and (b) α-helical proteins wrt the average structure. DSSP Plots of (c) β-hairpin H14 and (d) α-helical protein A.

atoms in the peptides were included in the analysis. The RMSD of the β-hairpin peptides is slightly higher than that of the α-helical proteins, which can be attributed to their direct solvent exposure leading and the resulting increased mobility. Nevertheless, in all systems the RMSD remains below 5 Å, indicating a strong conservation of the secondary structural motifs.

All β-hairpin peptides have a bend in the middle of the structure between residues 9 and 11 and extended β-sheets from residues 3 to 7 and 12 to 17, as shown exemplarily for β-hairpin peptide A in the DSSP plot in Fig. 3.5(c). These structures are consistent with the expectations of Ref. [99]. Fig. 3.5(d) shows exemplarily the DSSP plot of α-helical protein A, illustrating the three-helical α-helical secondary structure determined by CD and NMR spectroscopy.[103] The structure is stable along the 200 ns simulation.

Classical MD simulations with one tyrosine simulated as radical tyrosine were performed and, similar to the results above, the overall secondary structure was conserved as well. These simulations were performed to produce suitable initial structures for the QM/MM simulations, in which the residues involved in the PCET are close enough for the reaction to take place. The distances between the proton donor and acceptor atoms (nitrogen in histidine and tryptophan, oxygen in tyrosine) were measured. Structures in which this distance corresponded to a direct contact (≈ 3 Å) were used as initial structures for subsequent QM/MM simulations (see Fig. 3.6).



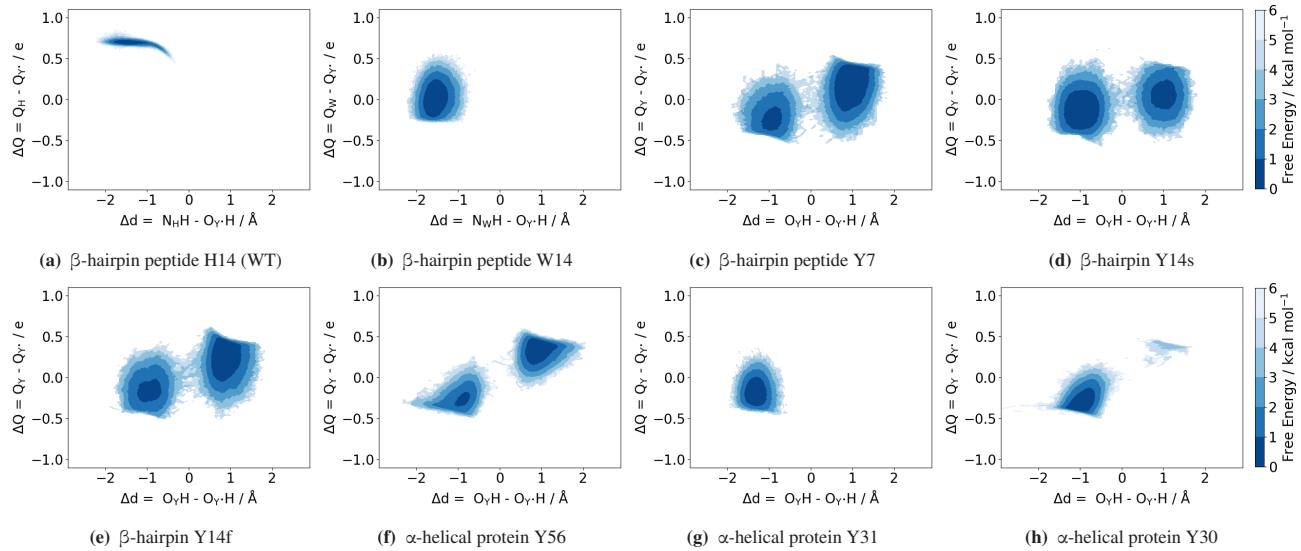
(a) Y14-s **(b)** Y56

Figure 3.6: Distance between oxygen atoms from classical simulations with one tyrosine in radical state, and zoom into structure with the minimal observed distance (≈ 3 Å) that served as input for the QM/MM simulation, shown exemplarily for (a) Y14s and (b) Y56.

3.3.2 Unbiased QM/MM Simulations

Unbiased DFTB3/MM simulations, i.e., simulations without enhanced sampling methods based on biasing potentials, were performed, including a radical tyrosine Tyr$^\bullet$ and a second protonated residue in the QM zone. The FESs are shown in Fig. 3.7.

No proton transfer was observed during the simulation time of 1 ns in the β-hairpin peptides H14 (WT) (Fig. 3.7(a)) and W14 (Fig. 3.7(b)), in which the radical tyrosine is expected to react with a histidine or with a tryptophan.[3,139,145] A narrow free energy minimum is observed at 0.7 e for the β-hairpin peptide H14 (WT). Indeed, the histidine is doubly protonated at the starting point of simulation (positively charged QM system), the radical stabilizes on the tyrosine residue, and so a formal hydrogen atom transfer appears highly unlikely. The partial charge of the hydrogen being transferred is $q_H = 0.3$ e, and the

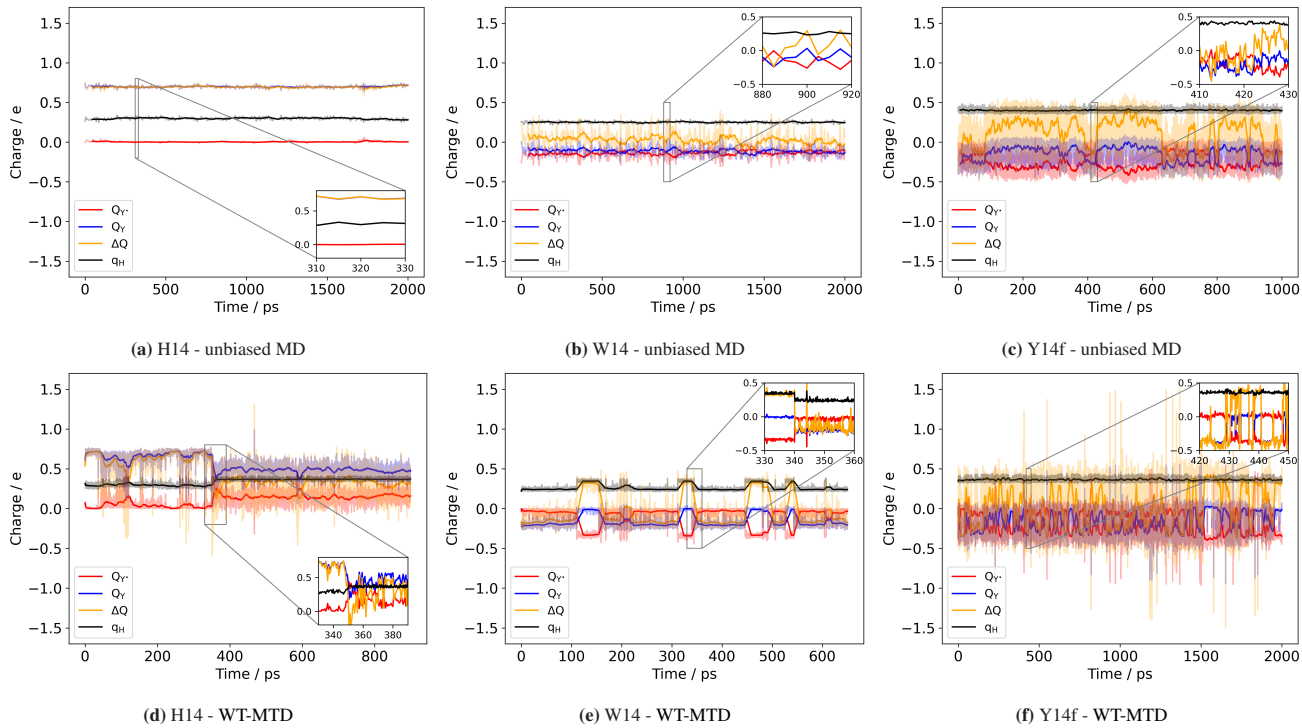


(a) β-hairpin peptide H14 (WT) **(b)** β-hairpin peptide W14 **(c)** β-hairpin peptide Y7 **(d)** β-hairpin Y14s

(e) β-hairpin Y14f **(f)** α-helical protein Y56 **(g)** α-helical protein Y31 **(h)** α-helical protein Y30

Figure 3.7: FESs of unbiased DFTB3/MM simulations of 1 ns for the different biomimetic peptides. The proton transfer CV is on the x-axis and the electron transfer on the y-axis. Contour lines are drawn every 1 kcal mol^{-1}. ΔQ is directly obtained from the Mulliken charges computed along the simulation, using the DFTB3/3OB method with spin polarization.

summed charge of the radical tyrosine is $Q_{\mathrm{Y}\bullet} = 0$ along the simulation, which results in $\Delta Q = Q_{\mathrm{H}} - 0 = 0.7\,\mathrm{e}$; see Fig. 3.8(a) for details.



(a) H14 - unbiased MD **(b)** W14 - unbiased MD **(c)** Y14f - unbiased MD

(d) H14 - WT-MTD **(e)** W14 - WT-MTD **(f)** Y14f - WT-MTD

Figure 3.8: Partial charges of H14, W14, and Y14f, shown as representative examples to illustrate the charge distribution on the molecules. Upper panels (a)–(c): unbiased MD simulations. Lower panels (d)–(f): selected walkers from MTD simulations in which the proton transfer was biased. The insets show magnified views of characteristic transfer regions to better track changes in the partial charges. Walkers were selected such that at least one clear transfer event was observed.

In the β-hairpin peptide W14, the proton remains on the tryptophan. On average, a charge of 0.15 e is localized on the tyrosine (Fig. 3.8(b)), as also reflected by the average value of ΔQ about 0.1 e. This delocalization may be due to the π-stacking between the phenol and indole rings.

Several proton transfer events occurred within the time scale of the simulations in both β-hairpin peptide Y14 configurations (stacked as in Fig. 3.7(d) and flipped as in Fig. 3.7(e)) and in the β-hairpin peptide Y7 (as in Fig. 3.7(c)). All of these three systems exhibit a similar rate-limiting barrier $\Delta G^{\ddagger}$ lower than 5 kcal mol^{-1}. We observe a strong fluctuation of ΔQ, which is also reflected by the free energy minima that are broad along the y-axis. During simulations, we observe a delocalization of the electron, which is more pronounced in the stacked geometry (Y14s) than in the flipped conformations Y14f and β-hairpin Y7, which are almost identical.

Only one or two proton transfer events happen during the unbiased simulations of the α-helical proteins Y56 (Fig. 3.7(f)) or Y30 (Fig. 3.7(h)) respectively. This led to a poor sampling of the transition region, and thus a high statistical uncertainty in the rate-limiting barrier $\Delta G^{\ddagger} \approx 5$ kcal mol^{-1}, slightly higher than for the above-mentioned β-hairpin peptides. No proton transfer reaction was observed in the α-helical protein Y31 (Fig. 3.7(g)). Overall, less electron delocalization appeared in Y56 and Y30 than in the β-hairpins. This observation can be related to the relative orientations of the tyrosine rings, as these were not in a π-stacked geometry during much of the simulation time (see Table 3.2). Y31 exhibits a similar propensity to π-stacking as the β-hairpin peptides Y7 and Y14f, and the reactant basin in the FES also appears similar to those in Y7 and Y14f.

Table 3.2: π-stacking interactions between aromatic residues involved in the PCET reaction displayed by relative occurrence. Thresholds and conditions were adapted from Ref. [146].

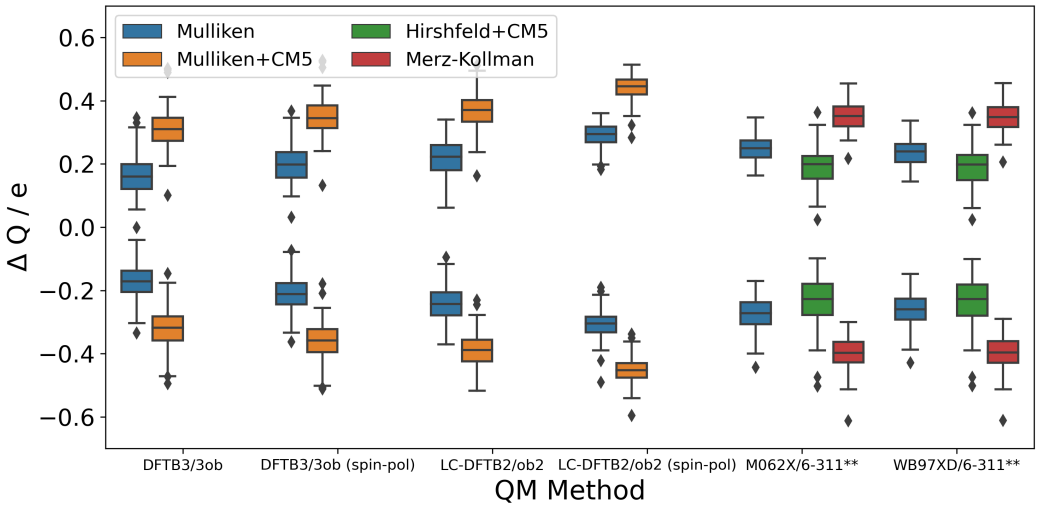
		β-H14	β-W14	β-Y7	β-Y14s	β-Y14f	α-Y56	α-Y31	α-Y30
free MD	π-stacked (%)	2.49	69.20	12.02	80.88	13.38	4.19	14.58	0.00
PT-Metad	π-stacked (%)	5.85	52.74	12.48	72.10	24.08	0.00	7.35	0.82

The QM method of our choice, DFTB3, tends to overdelocalize the electron density due to the self interaction error (SIE) of the underlying exchange-correlation functional of the generalized gradient approximation type.[61,63,147–149] However, overdelocalization of electron density is problematic when describing ET phenomena, and an example of a fine effect in which the overdelocalization manifests itself is the large width of the minimum-energy basins along the ET coordinate in the β-hairpin peptides (see Figs. 3.7(b)–3.7(e)).

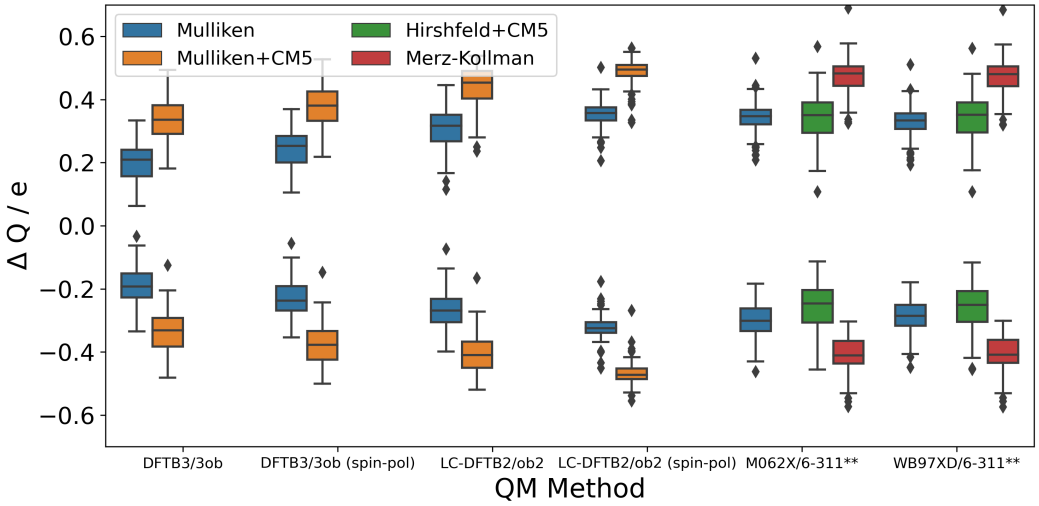
Note that while such an overdelocalization can occur in on-the-fly simulations, a large class of calculations used in previous work by others considers the electron to be constrained to one of the moieties,[150] effectively mitigating the overdelocalization problem. In this work, a comparison to several other DFT approaches was conducted to assess the accuracy of the description of ET with DFTB3.

As we can see in Fig. 3.9, the electron overdelocalization in DFTB3 calculations is largely resolved by employing LC-DFTB2, which was in a reasonable agreement with the two DFT reference methods (M06-2X and ωB97X-D).

With both of the DFT functionals considered, ΔQ obtained from MK charges are markedly larger in magnitude than those from CM5-corrected Hirshfeld charges (which only slightly differ from ΔQ yielded by Mulliken charges). While it appears difficult to decide which of those approaches is more accurate, we observe that ΔQ obtained using the Mulliken charges from LC-DFTB2 with spin-polarization are always within the interval spanned by those two extremes. Therefore, in this work, ΔQ from Mulliken charges obtained from LC-DFTB2 with spin-polarization is considered as a reasonable compromise, suffering little from electron overdelocalization due to SIE and thus suitable to describe ET phenomena.



(a) β-Y14s



(b) β-Y14f

Figure 3.9: Benchmark of charge analysis using various QM methods and partial charge schemes: DFTB3/3OB and LC-DFTB2/ob2 (each with and without spin polarization), as well as the DFT-methods M06-2X/6-311** and ωB97X-D/6-311**. For the DFTB methods, Mulliken charges and their CM5-corrected counterparts are used. For the DFT methods, both Mulliken and CM5-corrected Hirshfeld charges are evaluated, along with Merz-Kollman partial charges.

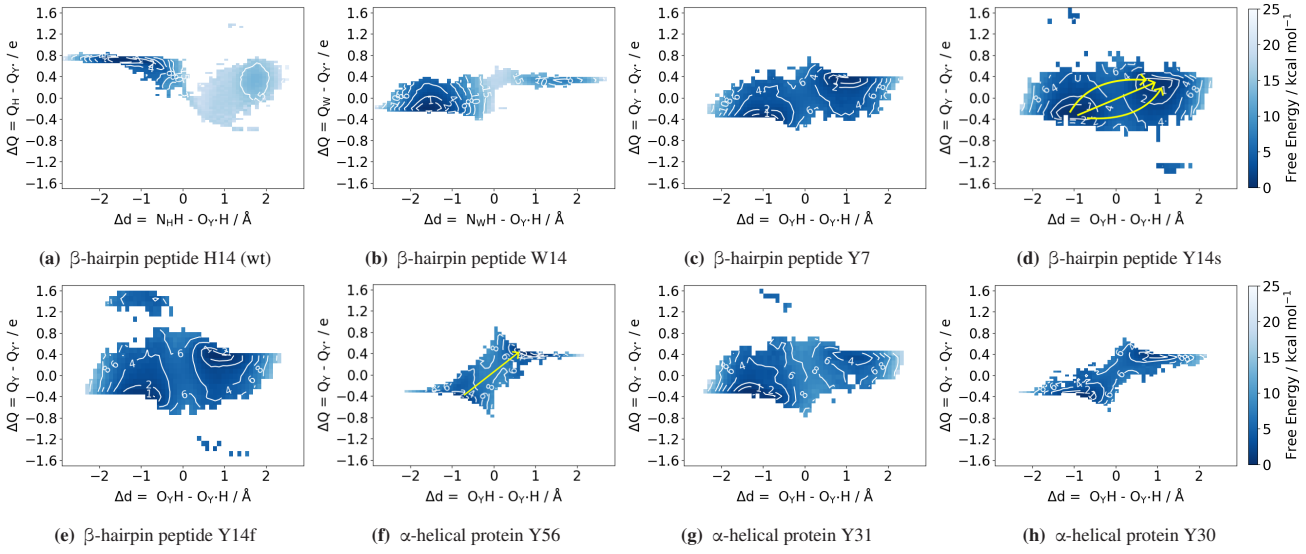
3.3.3 Metadynamics Simulations

One-dimensional (1D) MTD simulations were performed with a bias applied to the PT CV (s_{PT}) to better sample the transition region and the entire FES. In order to improve upon the overdelocalization of the electron density with DFTB3, which leads to an inappropriate description of the atomic charges during the PT mechanism, the Mulliken charges were recalculated with LC-DFTB2 along the MTD trajectories, prior to generating the FES. The 2D FES for PCET in all of the biomimetic peptides and proteins under investigation are shown in Fig. 3.10, and a concise overview in tabular form is provided in Table 3.3.

Although PCET is observed in all considered peptides and proteins, the free energy barrier is significantly higher whenever PCET occurs between a tyrosine and a different amino acid (histidine or tryptophan) than in the cases involving a pair of tyrosines. Indeed, the activation free energy $\Delta G^{\ddagger}$ exceeds 15 kcal mol^{-1} for the β-hairpin peptides H14 (wt) and W14 (Figs. 3.10(a) & 3.10(b)), while they are around 4 kcal mol^{-1} for the β-hairpin peptides containing a pair of tyrosines (Figs. 3.10(c)–3.10(e)) and between 4–8 kcal mol^{-1} in the α-helical proteins (Figs. 3.10(f)–3.10(h)). The reaction free energy

Table 3.3: Overview of the biomimetic peptides and their properties: Percentages of simulation time spent in a π-stack orientations (see Table 3.2 for details); free energy barrier $\Delta G^{\ddagger}$ and driving force of the reaction $\Delta G_{\mathrm{R}\to\mathrm{P}}$; total number of hydrogen bonds and character of the molecular environment of the PCET-active residues (a display of the environment can be found in Fig. 4.16).

name	amino acids Tyr$^{\bullet}$ and	π-stack / %	$\Delta G^{\ddagger}$ / kcal mol^{-1}	$\Delta G_{\mathrm{R}\to\mathrm{P}}$/ kcal mol^{-1}	H-bonds	environment
β-H14	His	6	15.2	12.4	2.03	solvent
β-W14	Trp	52	17.3	7.8	0.54	solvent
β-Y7	Tyr	13	4.6	—	2.14	solvent
β-Y14s	Tyr	72	4.0	—	2.39	solvent
β-Y14f	Tyr	24	4.7	−0.6	2.34	solvent
α-Y56	Tyr	0	8.1	1.2	1.28	protein
α-Y31	Tyr	7	7.5	2.3	1.94	solvent/protein
α-Y30	Tyr	1	5.1	1.0	1.27	protein



(a) β-hairpin peptide H14 (wt) **(b)** β-hairpin peptide W14 **(c)** β-hairpin peptide Y7 **(d)** β-hairpin peptide Y14s

(e) β-hairpin peptide Y14f **(f)** α-helical protein Y56 **(g)** α-helical protein Y31 **(h)** α-helical protein Y30

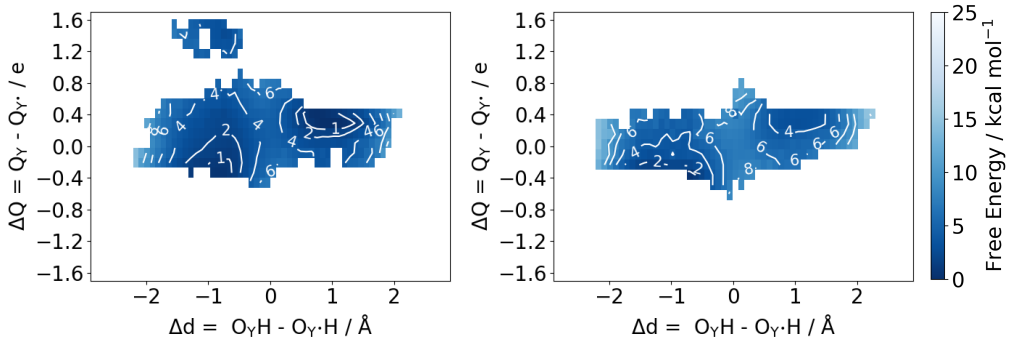
Figure 3.10: FESs of PCET in all tested systems using DFTB3/MM 1D MTD simulations with a bias on the PT CV. ET CV was corrected at LC-DFTB2/MM level. Contour lines are drawn at 2, 4, 6, 8, 10, and 15 kcal mol^{-1}. Arrows in (d) and (f) indicate the different appearances of the concerted mechanism of PCET between the Tyr side chains.

$\Delta G_{\mathrm{R}\to\mathrm{P}}$ of the reaction is close to zero for the Tyr$^{\bullet}$–Tyr systems, but it is quite high for the Tyr$^{\bullet}$–His/Trp systems. This first result indicates that the radical tyrosine state is favored, which is the reactant state in our simulations.

Additionally, PCET in two selected systems, β-Y7 and α-Y30, was investigated using the ff19SB force field for the biomolecule together with the optimal point charge (OPC) water model, which is the recommended combination.[120] The central regions in the resulting FES (see Fig. 3.11) are largely similar to those obtained with simple point-charge (SPC) water.

3.3.3.1 β-hairpin Peptides

In agreement with the unbiased simulations, a narrow minimum is observed at $\Delta Q = 0.7$ e and $\Delta d < 0$ in the FES of β-hairpin H14. It corresponds to a doubly protonated Histidine (note that the QM region carries a positive charge in this single case). This reactant basin is much deeper than the product basin ($\Delta G_{\mathrm{R}\to\mathrm{P}} = 12$ kcal mol^{-1}). The summed partial charges for the radical Tyr fluctuate around zero in this state, and the excess electron is located on the oxygen atom of the Tyr. Along the PT coordinate, the ΔQ coordinate fluctuates from 0 to 0.7 e for Δd values ranging from −1 to 1 Å, i.e., even when the proton still rests on the histidine (see Fig. 3.10(a)). As soon as the proton has transferred from His to Tyr, the charge of the histidine drops to $Q_{\mathrm{H}} \approx 0.4$ e, while that of the tyrosine increases to $Q_{\mathrm{Y}} \approx 0.2$ e; this excludes the charge of the transferred hydrogen



(a) Y7 **(b)** Y31

Figure 3.11: FESs of PCET in systems (a) Y7 and (b) Y31 simulated with OPC water instead of SPC using DFTB3/MM 1D MTD simulations with a bias on the PT CV. The central regions are similar to those obtained in SPC water.

q_H, which increases slightly (see Fig. 3.8(e)). This evolution of the charges is in agreement with a pronounced delocalization of the excess electron over both side-chains during the PT.

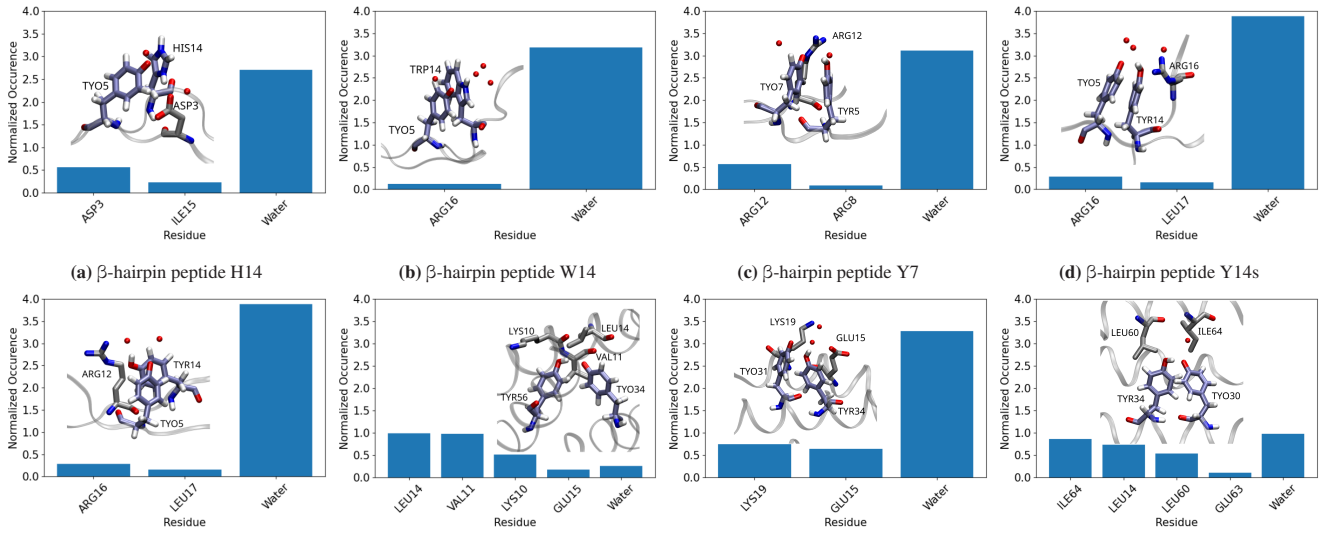
Furthermore, the minimum energy pathway between the reactant and the product states can be read from the FES. It appears that a partial ET can occur during the PT while the free energy increases up to the saddle point, which stands around 0.4 Å. A PT pathway seems as probable as a pathway combining a partial charge transfer at $\Delta d = 0$. The product basin suggests a delocalization of the electron charge, but histidine seems to be unable to form a stable radical in this state. The formal reaction mechanism can be described as a PT with the formation of an unstable radical tyrosyl cation. The proximity of Asp3 helps stabilizing the positive charge of the reactive center and it forms hydrogen bonds with His14, see Table 3.4, which also contributes to the stabilization of positively charged histidine. It may strengthen the asymmetry of this system.

Table 3.4: Average number of hydrogen bonds of atoms involved in the PCET reaction either to water molecules (solvent), to other amino acids residues (protein), or between the partners in the PCET reaction ($Tyr^{\bullet}$-X, with X is His in β-H14, Trp in β-W14, and Tyr in all other systems). Criteria to be met for hydrogen bonding: (i) $\text{angle}_{\text{H-Donor}\cdots\text{Acceptor}} \leq 30°$ and (ii) $\text{distance}_{\text{Donor}\cdots\text{Acceptor}} \leq 3.5$ Å.

	β-H14	β-W14	β-Y7	β-Y14s	β-Y14f	α-Y56	α-Y31	α-Y30
Solvent	0.97	0.52	1.43	1.97	1.67	0.16	1.59	0.56
Protein	0.82	0.00	0.30	0.06	0.20	0.25	0.07	0.08
$Tyr^{\bullet}$-X	0.25	0.02	0.41	0.36	0.47	0.87	0.28	0.63
total	2.04	0.54	2.14	2.39	2.34	1.28	1.94	1.27

In contrast, for β-hairpin W14, the reaction mechanism can be described as a formal hydrogen transfer with the ET CV changing from −0.2 e at the reactants to 0.4 e at the products (Fig. 3.10(b)) with a barrier of 17 kcal mol^{-1}. The reactant state with the tyrosyl radical lies 8 kcal mol^{-1} below the product state. Consequently, the deprotonated Trp seems less favorable than the radical Tyr, but still stable at 300 K with a backward free energy barrier of 9 kcal mol^{-1}. An electron delocalization between the two residues is observed when tryptophan is protonated, with a broader basin along the ΔQ coordinate (over 0.5 e), but also at the transition state to some extent. However, the electron localizes more on the tryptophan after PT to tyrosine. The saddle point in the W14 system is located at $\Delta d = 0.18$ Å, also underlining the asymmetric character of this system. Compared to the other β-hairpin peptides, the hydrogen bond network around the Tyr-Trp pair is relatively limited (see Table 3.4). However, the Tyr–Trp pair maintains a π-stacked orientation for more than half of the simulation time (Tab. 3.3), which can favor partial ET between the aromatic rings. Stable cationic and neutral tryptophan radicals have been characterized in proteins,[151–153] including Trp–Tyr pairs involved in the electron transfer chains in cytochrome c peroxidase,[154] DNA photolyase[143] and animal-like cryptochrome.[142]

For all three β-hairpin peptides with Tyr–Tyr pairs (Y7, Y14s, and Y14f), the central parts of the FES are similar, exhibiting a center of symmetry in agreement with the identical nature and solvent exposure of the respective PCET donors and acceptors. Moreover, the free energy barrier is 4 times lower than the previous asymmetric transfers, while the transition region is



(a) β-hairpin peptide H14 **(b)** β-hairpin peptide W14 **(c)** β-hairpin peptide Y7 **(d)** β-hairpin peptide Y14s

(e) β-hairpin peptide Y14f **(f)** α-helical protein Y56 **(g)** α-helical protein Y31 **(h)** α-helical protein Y30

Figure 3.12: Bar plots with normalized occurrence of amino acids and water in 3 Å distance of the atoms involved in PCET mechanism for all systems. Only residues with an occurrence ≥ 0.05 are displayed.

centered at $\Delta d = 0$ and $\Delta Q = 0$ for all three peptides. The product and reactant basins are relatively broad, suggesting energy-free partial electronic transfers during the proton motion. The larger occurrence of π-stacking conformations in β-hairpin Y14s (Tab. 3.3) is associated with a slightly flatter surface. A small preference for the neutral state of the tyrosines emerges as the minima are located at $\Delta Q = \pm 0.38$ e (Y7 and Y14f) or $\Delta Q = \pm\, 0.32$ e, which compensates the charge of the transferred proton. The surface's shape allows fully concerted or asynchronous transfer with a partial electron transfer up to 0.5 e before or after proton motion, which is indicated in Fig. 3.10(d) with three yellow arrows.

Additional basins at $\Delta Q = \pm 1.3$ e are observed on the FES for Y14s and Y14f. These correspond to intermediate states resulting from elementary ET events not accompanied by a PT, as they would appear in a step-wise PCET mechanism. Here, these are $\mathrm{Tyr}^+/\mathrm{Tyr}^-$ states, as opposed to the states that have both Tyr uncharged, which are located around $\Delta Q = \mp 0.4$ e (see above). Importantly, this means that the ET CV ΔQ, in combination with Δd as the PT CV, is in fact capable of distinguishing the four possible relevant states of the system.

3.3.3.2 α-helical Proteins

The three α-helical proteins differ in their relative positions in the protein chain: in Y30 and Y31, the tyrosines involved in the PCET reaction are on the same strand, whereas they are on opposite strands in Y56. Their close environment is also different: in Y30 and Y56, the reaction center is buried within the protein structure, while the Tyr side-chains in Y31 are rotated towards the solvent. Consequently, the tyrosines in Y56 and Y30 show only a small percentage of hydrogen bonds with nearby water molecules; rather, they are hydrogen-bonded with each other for more than half of the trajectory time (Tab. 3.4). No π-stacking is observed between them (Tab. 3.3). The reaction center in Y56 and Y30 proteins is constantly in contact with the neighboring amino acids, such as Leu14 and Val11 in Y56, and Ile64 and Leu60 in Y30 (see Fig. 3.12). In contrast, the solvent exposure in Y31 is similar to β-hairpin Y14 (see Table 3.4).

The accessible region on the FES of the α-helical protein Y31 (Fig. 3.10(g)) is spread along the ΔQ-axis much like in the β-hairpins Tyr–Tyr FES. Similarly, the two minima in the FES are at $\Delta Q = -0.31$ e and $\Delta Q = 0.36$ e, close to the positions of the minima on those surfaces described above. However, the reactant state is 3 kcal mol^{-1} more favorable than the product state, reflecting a slight asymmetry of the PCET, and the free energy barrier $\Delta G^{\ddagger}$ is of 7 kcal mol^{-1}, so 3 kcal mol^{-1} higher than in Y7 or Y14. Y31 is the only peptide with two tyrosines which features an asymmetric FES. A possible cause of this is the rather different character of the molecular environment: there is a strong salt bridge Lys19–Glu15 right next to Tyr34, and

the strong induced electric field may lead to the preference of protonated Tyr34. The saddle point is shifted to $\Delta d = 0.09$ Å, in agreement with the asymmetry character of the system. Apparently, the FES for Y31 supports both synchronous PCET and asynchronous with a partial PT. Also, an additional basin on the FES is observed around $\Delta d = -1$ Å and $\Delta Q = +1.4$ e. Similarly to the FES of β-hairpin peptides Y14s and Y14f, this corresponds to an intermediate state resulting from an ET step that has not been accompanied by a PT. The basin itself is not well resolved because of the lack of sampling of this electronic state.

The α-helical proteins Y56 and Y30 exhibit a nearly symmetrical FES with a driving force of 1 kcal mol^{-1} (Figs. 3.10(f) & 3.10(h)). Their shape is similar, as well as the location of the minima at $\Delta Q = -0.37$ e and 0.36 e, in relatively narrow basins. Both surfaces also differ from other FES in their more restricted accessible range of ΔQ values: in Y30, ΔQ can fluctuate over ca. 0.5 e for a given position of the proton, while it can reach 0.8–1 e for Y31. Then, we observe a narrow corridor connecting the reactant and the product. Consequently, only a concerted transfer can be drawn on the FES in these two systems (see yellow arrow in Fig. 3.10(f)). However, their free energy barriers differ by 2 kcal mol^{-1}, from 4.6 kcal mol^{-1} for Y30 to 6.6 kcal mol^{-1} for Y56. This difference may arise from variations in conformation or the local environment. Visual inspection reveals that in Y30, the aromatic rings are closer together and a partial overlap of their π-system can occur during the simulation. On the contrary, in Y56, the rings adopt a flipped geometry and stay at a larger distance (see Fig. 3.13). The proximity of the aromatic rings in combination with the conformational arrangement that facilitates the ET motion in Y30 through the π-system could explain the increase in the reaction rate. Besides, a water molecule approaches the reaction center of Y30, but it only does so rarely in Y56 (Fig. 3.12).

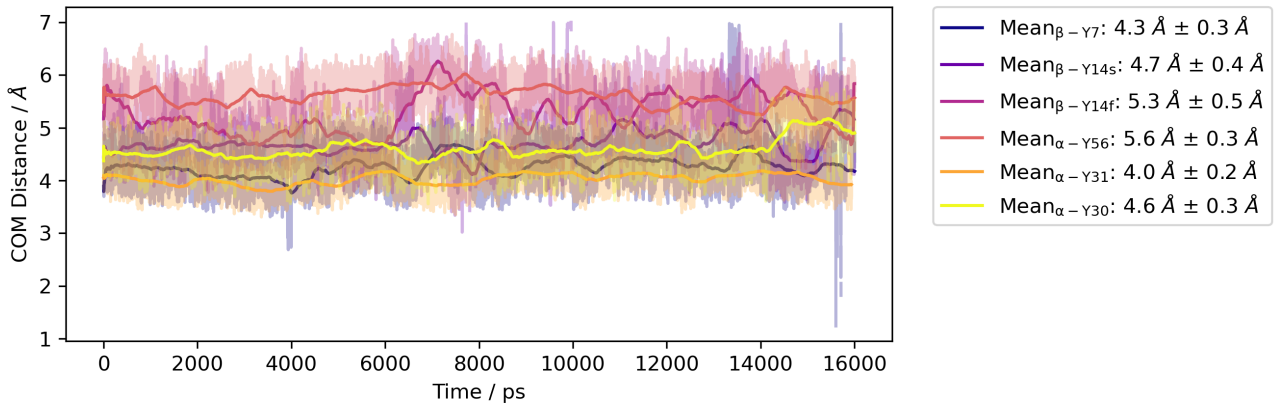


Figure 3.13: Distance between center of mass (COM) of tyrosine residues involved in PCET.

3.4 Discussion and Conclusions

The application of the efficient density-functional method DFTB3 in a QM/MM framework allowed the generation of long MD trajectories for several biomimetic peptides in which PCET reactions take place. By combining our setup with an enhanced sampling scheme, extended regions of the conformational space are covered, and even high-energy barriers are overcome within the available simulation time.

Our current protocol only makes it possible to sample the most stable electronic state along the PT coordinate, so to speak, an adiabatic (ground) electronic surface. Any higher-lying electronic states are not explored, and the coupling between them cannot be evaluated. Consequently, on the basis of our current data, no decision can be made on the (non-)adiabatic nature of the ET, nor on the ET rate. Another protocol dealing with the specificity of ET, such as the coupling and the energy gap between the different involved charge state must be developed to answer these issues.

Further, there are two important considerations that any computational study of ET needs to account for. One is the degree of spatial extension of the electron being transferred. It is especially DFT-based approaches that tend to overly delocalize the electron density, and DFTB is no exception. In this work, the use of LC-DFTB2 largely avoids overdelocalization of excess electron, which would otherwise be observed with DFTB, and improves the description of ET.

The other is the CV that shall describe the electronic behavior. In this work, the difference of summed atomic charges ΔQ turned out to be an effective CV that makes it possible to distinguish all of the states of the system that are relevant for PCET. The current simulation protocol evaluates ΔQ only after the simulation has been performed, instead of placing a biasing potential on ΔQ in what would be a 2D MTD procedure. The latter approach would clearly be desirable, and the aim of our ongoing work is the optimization of the CP-DFTB framework that would make it practicable for realistic biomolecular complexes. Still, the fact that the ET-associated CV is not biased results in limited sampling: while two charge states resulting from elementary ET events are observed for some of the systems under investigation in the PT product or reactants state, they are not sampled along the PT reaction path. It appears that fully converged 2D FES will only become available with the full 2D MTD scheme that considers biasing potentials as a function of both PT and ET CV; this will also make it possible to deduce whether a concerted or step-wise mechanism is in action in adiabatic PCET.

Our free energy barriers to direct PCET taking place between two Tyr side-chains range between 4–7 $\mathrm{kcal\,mol^{-1}}$, in good agreement with these previous reports.[17,111,112] Furthermore, our value of $\Delta G^{\ddagger}_{\mathrm{Tyr-Trp}}$=17 $\mathrm{kcal\,mol^{-1}}$ for the peptide β-W14 lies within the above mentioned range of $E_{\mathrm{a}} = 7-19\,\mathrm{kcal\,mol^{-1}}$.[110] Our results are thus in quantitative agreement with previous studies. Inspecting the six peptides and proteins featuring PCET between two Tyr side chains, we observe two different kinds of transition region topology. For instance, the α-helical proteins Y56 and Y30 FES describe a relatively narrow transition state area, which suggests synchronous ET and PT. On the contrary, the transition region is relatively broad in the FESs of the other peptides and proteins under investigation, as exemplified by the case of β-Y14s. In such a situation, different pathways can connect the reactant and product basins with an identical free energy cost. Although all of these pathways still correspond to a concerted, one-step PCET, they may involve a partial transfer of the electron taking place before or after PT. We call this process asynchronous concerted PCET and emphasize that it consists of a single-step mechanism and features no intermediates.

In summary, we find that the PCET mechanism is affected both by the orientation of the residues involved in the reaction and by environmental factors. Specifically, it was revealed how the interactions between the reaction center and the nearby protein components, as well as solvent exposure, affect the mechanism and kinetics of the PCET. We take advantage of the low computational cost of our approach for a better sampling of this environment and different conformations of the reactive center.

Our development so far has focused on increased efficiency and, thus, precision of the simulations by maximizing the sampling of configurational space. It can be readily adopted to investigate PCET occurring in more complex reaction centers, such as PCET processes involving three amino-acid side-chains. There is a possible participation of tryptophan, tyrosine,

and a nearby positioned histidine in a PCET reaction in animal-like cryptochrome and *Methanosarcina mazei* class II DNA photolyase,[142,143] as explained in chapter 4.

Another case of considerable research interest is water-mediated PCET, in which a larger number of water molecules may be involved. Such an application might utilize one of the previously developed CVs for description of long-range proton transfers, which were demonstrated to be practicable for long-range water-mediated PT processes in biomolecular complexes.[81]

In an ongoing work, we aim to implement an enhanced sampling workflow that includes biasing potentials on both the proton and electron transfer CVs,[60,66] as detailed in chapter 5. Further, we are aiming to improve the accuracy of the approach by passing to more refined DFTB models, based on LC-DFTB for a better description of the electronic behavior, but keeping a good description of PT and hydrogen bonding. A pertaining limitation is that the method, as it stands currently, is restricted to the ground electronic state, so that any non-adiabaticity and quantum effects, potentially affecting PCET in enzymes,[17,22] cannot be considered.

4 PCET Reactions in Photolyases and Cryptochromes

4.1 Introduction

The photolyase/cryptochrome (PL/CRY) family comprises blue-light-sensitive flavoproteins present in all kingdoms of life, from bacteria to mammals.[155,156] These proteins are involved in diverse biological functions, including deoxyribonucleic acid (DNA) repair, transcriptional regulation, magnetoreception, and circadian rhythm control. Most members of the PL/CRY family share two common features: sensitivity to blue light, mediated by two chromophores, and the ability to bind DNA.

Photolyases are primarily associated with light-driven DNA repair processes[157–159] while cryptochromes play roles in transcriptional regulation in plants,[160,161] algae[106,162] and various animals such as *Drosophila*,[163] zebrafish,[164] and particularly migratory birds. In the latter, cryptochromes are thought to mediate magnetoreception via a radical pair mechanism.[165–169]

Phylogenetic studies suggest that current proteins evolved from a common ancestor and diverged into multiple functional subgroups. Functional diversity has emerged through mutations during evolution, but following different ways. For instance, photolyases that repair cyclobutane pyrimidine dimers (CPD), a UV-induced DNA lesion, belong to different classes (I, II, and III), which in some cases are more closely related to cryptochromes than to each other. This raises the question of how the mutations tune the protein functions.[160,170–172]

One of the common actions of PL/CRY involves the photo-activation of the flavine adenine dinucleotide (FAD) cofactor, which is bound to the protein in a characteristic U-shaped orientation. Upon photoexcitation, FAD is reduced, typically facilitated by a conserved chain of aromatic residues, predominantly tryptophans. This tryptophan triad (or tetrad, in some cases) enables stepwise electron transfer (ET) to the FAD (see Fig. 4.1). In some proteins, a fourth tryptophan or a tyrosine residue, both located near the protein surface, acts as the terminal, intrinsic electron donor during photoactivation. Light is first absorbed by a primary light-harvesting chromophore, usually a folate or a flavin derivative, which transfers excitation energy to the earlier reduced $FADH^-$ via resonance energy transfer.[171] The resulting change in the redox state of the FAD is essential for protein function: in PLs, it enables photo-induced ET to the DNA lesion, triggering repair; in CRYs, it initiates conformational changes that launch signaling cascades or generate long-lived radical pairs that may participate in magnetoreception.

Multiple studies have investigated timescales and kinetics of the electron transfer (ET) chain, both experimentally[173–180] and theoretically.[143,181–183] In the latter, numerous studies are focusing on magnetoreception in cryptochromes[184–186] or DNA repair mechanisms in photolyases.[181,182,187] For a comprehensive overview, see Ref. [158]. Additionally, considerable interest has been devoted to the protonation of the semi-reduced anion radical ($FAD^{\bullet -}$).[188]

The class II photolyase from archaea *Methanosarcina mazei* (MmCPDII) is involved in the light-driven repair of the CPD. There is a cationic tryptophan radical WH^+ at the end of the chain and there is experimental evidence for a tyrosyl radical, leading to the proposal of a PCET between $W388H^{\bullet +}$ and Y345. Possible proton acceptors near $W388H^{\bullet +}$ are a histidine, an aspartate, and a glutamate. The latter was experimentally excluded as possible proton acceptor, as mutant studies revealed the same deprotonation kinetics in E378 as in the wild type (wt). One can also find experimental evidence for an ideally positioned water cluster that could facilitate deprotonation to bulk water or to a final proton acceptor at the protein surface.[143]

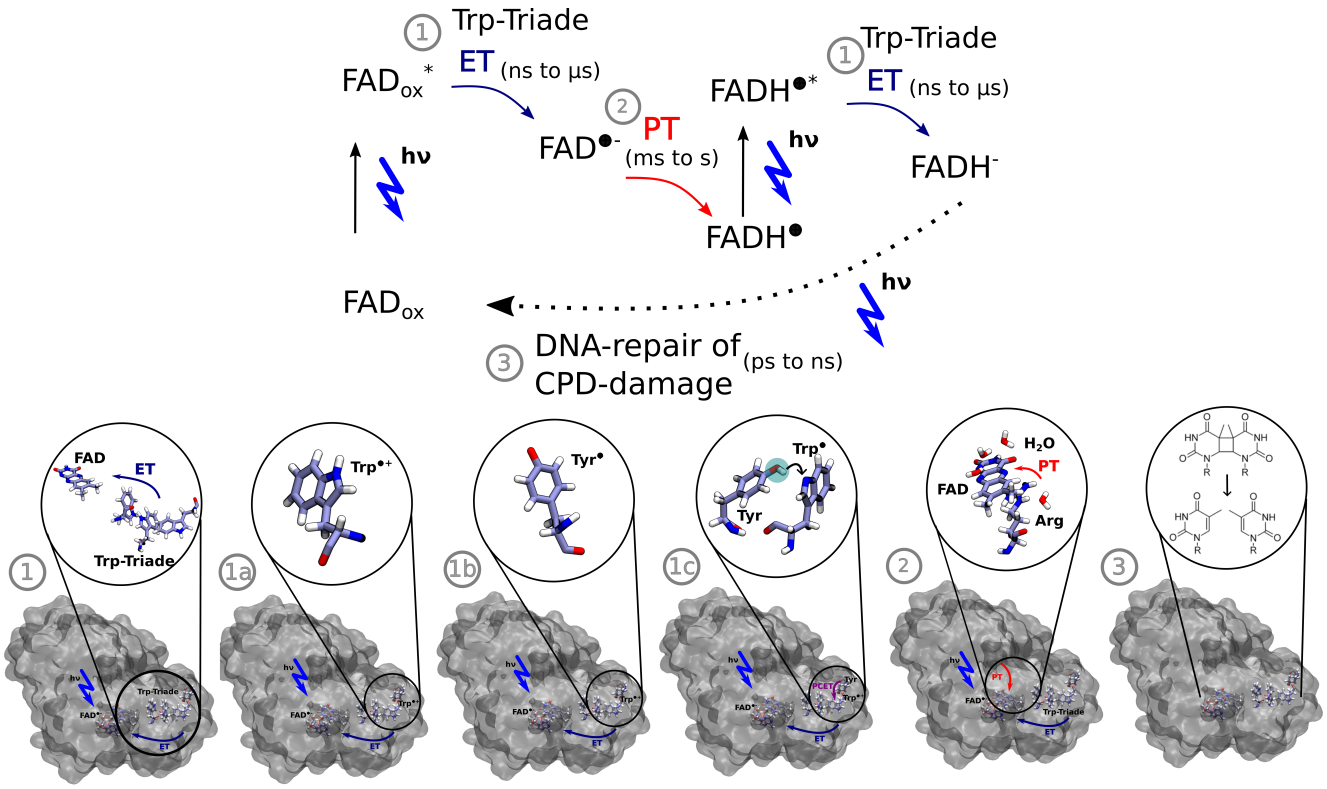


Figure 4.1: Photoactivation process in MmCPDII: FAD_{ox} is excited by blue light and reduced to $FAD^{\bullet -}$ via electron transfer from the Trp-triad (step **1**), generating a cationic tryptophan radical $WH^{\bullet +}$ (step **1a**). Experimental detection of a tyrosyl radical $Y^{\bullet}$ (step **1b**) suggests a proton-coupled electron transfer (PCET) between $W^{\bullet}$ and Y (step **1c**). $FAD^{\bullet -}$ is then protonated (step **2**), followed by further reduction from the electron transfer chain (step **1**), leading to light-triggered DNA-repair of the CPD lesion (step **3**).

In several cryptochromes, a fourth member of the Trp-triad is reported, or a nearby tyrosine at the end of the electron transfer chain.[106,142,169,176,180] The animal-like cryptochrome from green alga *Chlamydomonas reinhardtii* (CraCRY) functions as a blue and red light photoreceptor and modulates the expression of numerous genes. Photoreduction from the oxidized FAD to the neutral radical state is induced by blue light, followed by further reduction to the fully reduced state by red light.[189] For this protein Ref. [142] proposes the involvement of a nearby aspartate residue in the PCET mechanism, which may proceed via proton transfer followed by electron transfer (proton transfer (PT)/ET), electron transfer followed by proton transfer (ET/PT), or concerted PCET pathway.

In this work, the different possible PCET mechanisms are investigated using multiscale simulations combined with metadynamics to enhance sampling.[97] Our goal is to provide a comprehensive picture of the molecular processes in the two photoreactive proteins, MmCPDII and CraCRY, and to identify general principles that could guide protein design and deepen our understanding of biological PCET.

4.2 Methods

4.2.1 System Preparation

The protein structures used in this study are based on the X-ray crystallographic structures of MmCPDII (PDB ID: 7VJ1)[188] and CraCRY (5ZM0).[162] Missing residues not resolved in the crystal structures were modeled using the SWISS-Model server.[190] The H++ web server[191–194] was used to determine the protonation states of titratable amino acids. Both proteins were simulated using the Amber ff14SB force field and solvated with the SPC water model,[119,120] in addition to the crystallographic water molecules present in the original structures.

The atomic charges of the protonated radical tryptophan ($WH^{\bullet+}$) and the deprotonated, neutral tryptophan ($W^{\bullet}$) were obtained using the restrained electrostatic potential fitting method (RESP), as implemented in Antechamber,[121,122] part of the AmberTools software suite.[119] The underlying electron densities were calculated at the HF/6-31G* level of theory using Gaussian09.[123] In the same manner, parameters were generated for the three redox states of FAD: FAD_{ox}: flavoquinone, FAD: flavosemiquinone, and FADH: flavohydroquinone. Simulations were initially performed with both flavoquinone and flavosemiquinone to validate system stability; however, all subsequent QM/MM simulations were conducted using flavosemiquinone, as it represents the redox state of FAD following the first reduction by the tryptophan triad.

The simulation box was cubic, with a minimum solute-to-box distance of 12 Å, and a salt concentration of 0.1 mM was maintained by adding an appropriate number of Na^+ and Cl^- ions. Periodic boundary conditions were applied. The Verlet neighbor-searching algorithm was used, with a cutoff distance of 12 Å for both neighbor searching and Lennard-Jones interactions. The long-range electrostatic interactions were treated using the particle mesh Ewald (PME) method.

4.2.2 Classical Equilibration and Simulation

The simulation protocols followed those established in our previous publication on PCET in biomimetic peptides,[97] employing classical MD simulations for system preparation and to identify suitable starting structures for the PCET reaction. Energy minimization was carried out using the steepest descent algorithm until the maximum force fell below $1000\,kJ\,mol^{-1}\,nm^{-1}$. NVT equilibrations were performed at 300 K for 10 ns with a time step of 2 fs, using Bussi thermostat (v-rescale) for temperature coupling. All bonds involving hydrogen atoms were constrained using the LINCS algorithm. NPT equilibration was performed for 1–5 ns, depending on the system, with a time step of 2 fs. Pressure was maintained at 1.0 bar using the Parrinello-Rahman barostat. Production runs were conducted with the same parameters as the NPT equilibration, lasting between 30–250 ns, depending on the system. All classical simulations were performed using the GROMACS program package (version 2020).[124,125]

4.2.3 QM/MM Setup

All QM/MM simulations were performed using DFTB3 as quantum mechanical (QM) method,[58] as implemented in our local version of DFTB+ patched with GROMACS,[127,128] with the standard 3OB parameter set.[59] The accuracy of this method is comparable to that of full DFT calculation employing medium-sized basis sets.[58] The composition of the QM region varied depending on the residues involved in the proposed PCET reaction (see Table 4.1). To saturate the valence shells of the relevant Cα atoms, hydrogen link atoms were introduced in the QM regions at the Cα–Cβ covalent bonds crossing the QM–MM boundary. The spin-polarized formalism of DFTB was employed to describe the electronic structures with radical character.[126] The MM region was modeled using the Amber ff14SB force field.[195] Both the distance between the proton donor and acceptor atoms, and the sum of the distances between the transferring hydrogen and each donor/acceptor atom were restrained using an upper-wall restraint of 3 Å using the PLUMED plugin (version 2.5.1).[47,117] The QM/MM interface was realized through our local implementations in DFTB+ and GROMACS.[127,128]

Table 4.1: Overview of the reactions simulated in this work with state of the W radical and additional residues in the QM region as well as the parameters for the well-tempered metadynamics (WT-MTD) simulations.

	reaction	W state	PCET partner(s)	Charge	WT-MTD parameters			
					σ	bias factor	# of walkers	time / ns
MmCPDII	**1a**	W388H$^{\bullet+}$	D357^{-} + Y345	0	0.02	21	8	7.2
	1b	W388H$^{\bullet+}$	H356 + Y345	1	0.02	21	8	7.4
	1c	W388H$^{\bullet+}$	H356 + Y345	1	0.02	15	8	5.2
	2a	W388$^{\bullet}$	Y345	0	0.02	21	8	5
	2b	W388$^{\bullet}$	H356 + Y345	0	0.02	21	8	7
CraCRY	**1a**	W322H$^{\bullet+}$	D321^{-} + Y373	0	0.02	21	8	$\geq$6
	1b	W322H$^{\bullet+}$	D321^{-} + Y373	0	0.02	21	8	2.8
	2a	W322$^{\bullet}$	Y373	0	0.02	15	8	5.6

4.2.4 Metadynamics Simulations and Recalculation of ET CV

Well-tempered metadynamics (WT-MTD) simulations, in which the proton transfer (PT) was biased, were performed within a multiple-walker framework using a combination of GROMACS, density functional tight binding (DFTB)3, and PLUMED. The initial Gaussian height was set to 0.5 kJ mol^{-1} with a width of $\sigma = 0.02$ Å and a bias factor of $\gamma = 15$ or 21 depending on the system. Communication of the walkers' hill files occurred every 500 steps, and the PT collective variable (CV) grid ranged from -4 to 4 Å. The keyword CALC_RCT was activated in simulations where additionally the electron transfer (ET) was of interest; it directly provides the reweighted bias and the time-dependent factor $c(t)$.[44] The number of walkers was generally set to 8; any deviations from these parameters are listed in Table 4.1. All walkers were simulated with a time step of 0.5 fs for a total simulation time ranging from 1.3 to 4.5 ns, as required for the convergence of the free energy surfaces.

For reactions **1a**, **1b**, and **2b** in MmCPDII and reaction **1a** in CraCRY, we simulated a 2D-metadynamics simulation with the first CV being the initial proton transfer (PT1) and the second CV being the secondary proton transfer (PT2). Since PT2 likely corresponds to the PCET step, we recalculated the ET CV once the WT-MTD simulations were completed, following the protocol established in our previous study on PCET in biomimetic peptides.[97] Specifically, we performed single-point calculations using spin-polarized LC-DFTB[63] with the OB2 parameter set,[129] incorporating the protein environment in a hybrid quantum mechanics/molecular mechanics (QM/MM) fashion via the electrostatic potential (ESP). Matrix-based screening was chosen to reduce computational cost. With that procedure, we can capture the entire reaction (PT1 and PT2 together). For the remaining reactions, **1c** and **2a** in MmCPDII, and **1b** and **2a** in CraCRY, we biased the proton transfer along a metadynamics simulation and recalculated the ET CV to generate a two-dimensional free energy surface (FES).

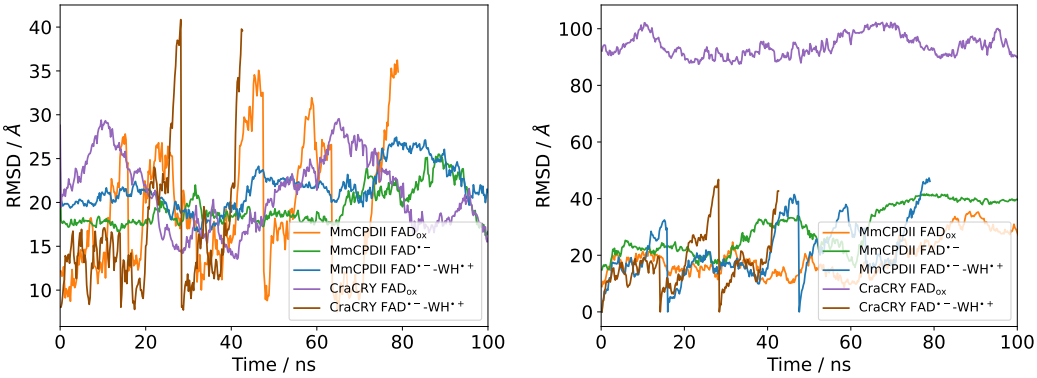
All data was processed and visualized with python[134] using the matplotlib,[135] NumPy,[136] pandas,[137] and seaborn[138] libraries.

4.3 Results and Discussion

4.3.1 Classical MD Simulations

We investigated the primary electron donor in the proteins MmCPDII and CraCRY, which is hypothesized to be a tyrosine, Tyr (Y) near the terminal tryptophan (W) of the electron transfer chain.

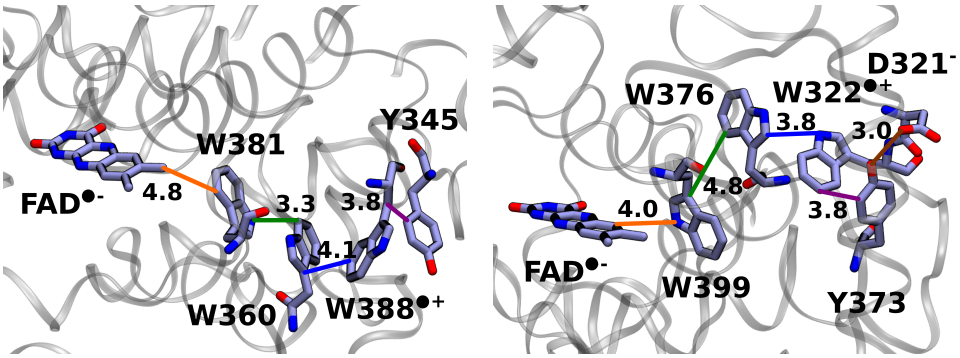
Classical molecular dynamics (MD) simulations were first carried out to assess the structural stability of both proteins. Fig. 4.2 shows the root mean square deviation (RMSD) values for different FAD oxidation states. Overall, comparison to the average structure proved to be more reliable than comparison to the initial structure, which corresponds to the end structure from the NPT equilibration step. This is especially true for the oxidized FAD state of CraCRY, which deviates significantly from its starting structure. Given the larger size of the proteins compared to the biomimetic peptides in chapter 3, the somewhat higher RMSD values remain acceptable.



(a) Average structure **(b)** Starting structure

Figure 4.2: Secondary structure conservation of proteins MmCPDII and CraCRY is evaluated with RMSD plots wrt (a) the average structure and (b) the end structure of the classical equilibration (NPT ensemble).

The electron transfer chains from the crystal structures are shown in Figs. 4.3(a) and 4.3(b) for MmCPDII and CraCRY, respectively, along with the key inter-residue distances. These distances were monitored throughout the classical MD simulations (Figs. 4.4(a) to 4.4(e)). In MmCPDII, the tryptophan triad distances remain close to the crystallography values, with moderate fluctuations. The largest variability is observed in the distance between the terminal tryptophan W388 and tyrosine Y345. For subsequent QM/MM simulations, representative structures were selected from the simulation with FAD in the



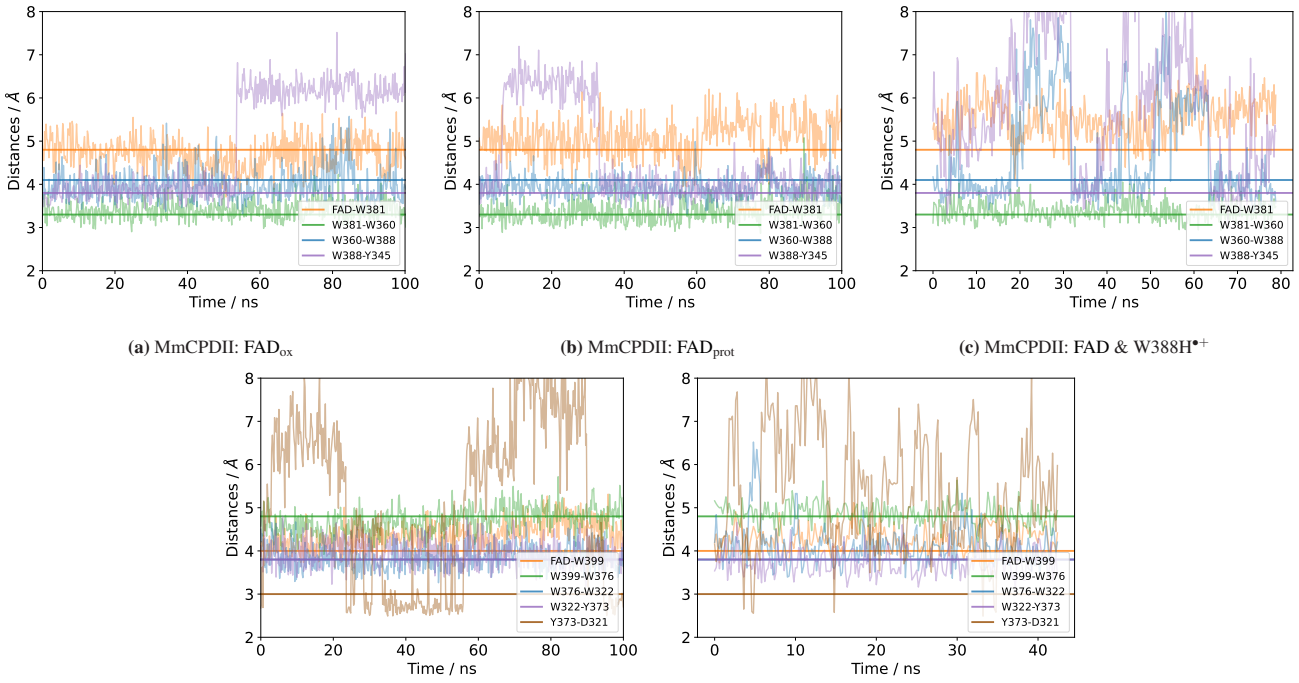
(a) MmCPDII (PDB 2XRZ) **(b)** CraCRY (PDB 5ZM0)

Figure 4.3: Crystallographic structure of the electron transfer chain of (a) MmCPDII and (b) CraCRY with important interatomic distances indicated between the involved residues: orange FAD$^{\bullet-}$-W1; green W1-W2; blue W2-W3; purple W3-Y; and brown Y-D (only in CraCRY).

flavosemiquinone state ($FAD^{\bullet}-$) and W388 in its cationic radical form ($W388^{\bullet+}$). Input structures were chosen such that the $W388^{\bullet+}$-Y345 distance corresponded to a direct contact (< 4 Å).

In CraCRY, not only the W322-Y737 distance is relevant, but also the distance between Y737 and the nearby aspartate (D) $D321^{-}$. While most other inter-residue distances remain stable at values close to the crystal structure, the Y373-$D321^{-}$ distance fluctuates strongly between 3 and 7 Å. For the subsequent QM/MM simulations, structures were chosen from the simulation with $FAD^{\bullet}-$ and $W388^{\bullet+}$, selecting configurations in which the Y373-$D321^{-}$ distance was approximately 3 Å.

In summary, the classical MD simulations confirm that both MmCPDII and CraCRY remain structurally stable across different FAD redox states. At the same time, they provided representative input structures with appropriate residue configurations for the subsequent QM/MM simulations of the PCET pathways.



(a) MmCPDII: FAD_{ox} **(b)** MmCPDII: FAD_{prot} **(c)** MmCPDII: FAD & $W388H^{\bullet+}$

(d) CraCRY: FAD_{ox} **(e)** CraCRY: FAD & $W388H^{\bullet+}$

Figure 4.4: Monitored distances of the key interatomic distances between residues in the electron transfer chain along the classical MD simulations. Crystallographic distances are indicated by solid lines. Different redox states for FAD and the terminal W in the electron transfer chain. Color codes follow as: orange $FAD^{\bullet-}$-W1; green W1-W2; blue W2-W3; purple W3-Y; and brown Y-D (only in CraCRY).

4.3.2 MmCPDII

There are several possible PCET reaction pathways in MmCPDII, as two potential proton acceptors are located near $W388H^{\bullet+}$ and Y345 (see Figure 4.5). The terminal tryptophan of the electron transfer chain, $W388H^{\bullet+}$, can transfer its proton either to $D357^{-}$ (**1a**) or H356 (**1b**). This is followed by a subsequent PCET between $W388H^{\bullet}$ and Y345. Alternatively, there could be PT from Y345 to H356 and simultaneous an ET between $W388H^{\bullet+}$ and $Y345^{-}$ (**1c**). If deprotonation of $W388H^{\bullet+}$ to bulk water or a final proton acceptor has occurred previously, then either a direct PCET between $W388^{\bullet}$ and Y345 is possible (**2a**), or again, there is PT between Y345 and H356 followed by a PCET reaction between $W388^{\bullet}$, $Y345^{-}$ and H356 (**2b**).

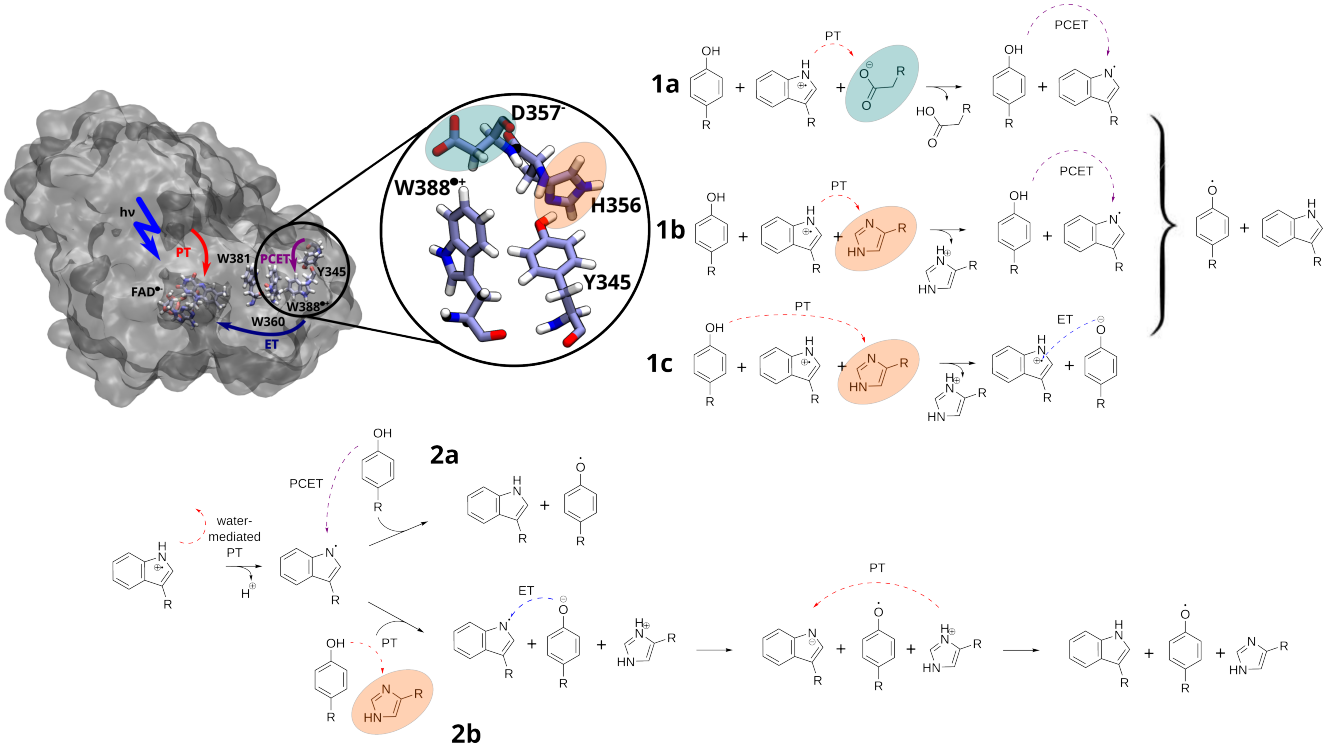
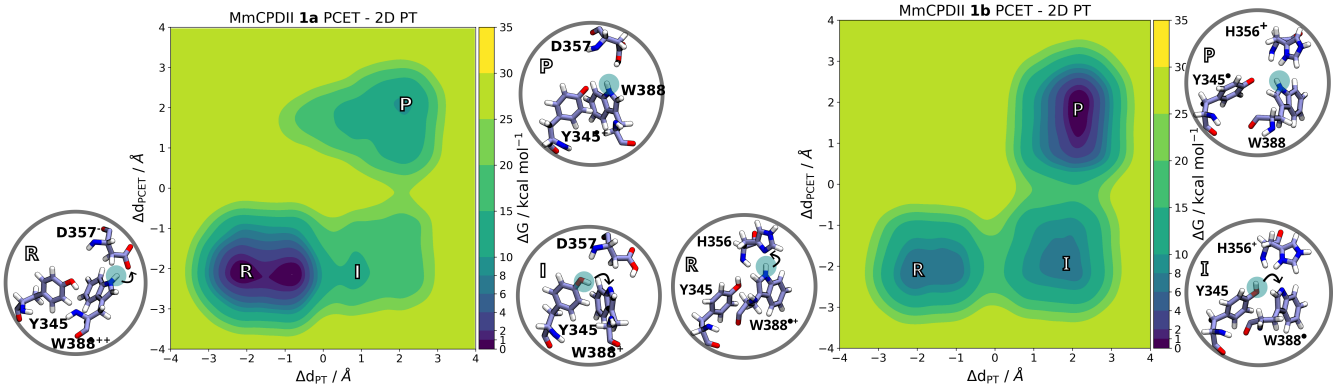


Figure 4.5: Overview of possible PCET reactions in MmCPDII: **1a** proton transfer from W388H$^{\bullet+}$ to D357^{-}, followed by PCET between W388$^{\bullet}$ and Y345; **1b** proton transfer from W388H$^{\bullet+}$ to H356, followed by PCET between W388$^{\bullet}$ and Y345; **1c** PCET with different acceptors, where Y345 donates a proton to H356 and an electron to W388H$^{\bullet+}$; **2a** direct PCET between W388$^{\bullet}$ and Y345 after prior proton transfer from W388H$^{\bullet+}$ to an external acceptor; **2b** PCET with different acceptors, where Y345 donates a proton to H356 and an electron to W388H$^{\bullet+}$ after prior proton transfer from W388H$^{\bullet+}$ to an external acceptor.

4.3.2.1 W388H$^{\bullet+}$: Cationic Radical Tryptophan in Reactant State

In MmCPDII, two possible proton acceptors are located near W388H$^{\bullet+}$ and Y345: either D357^{-} (which must be deprotonated for proton transfer), or singly protonated, neutral H356. Figures 4.6(a) and 4.6(b) show the FES obtained from two-dimensional WT-MTD simulations. The x-axis represents the proton transfer from W388H$^{\bullet+}$ to either D357^{-} or H356, while the y-axis corresponds to the secondary proton transfer from Y345 to W388$^{\bullet}$, presumably coupled to the electron transfer. In reaction **1a**, the reactant state "R" is lowest in energy. First, proton transfer from W388H$^{\bullet+}$ to D357^{-} occurs, forming



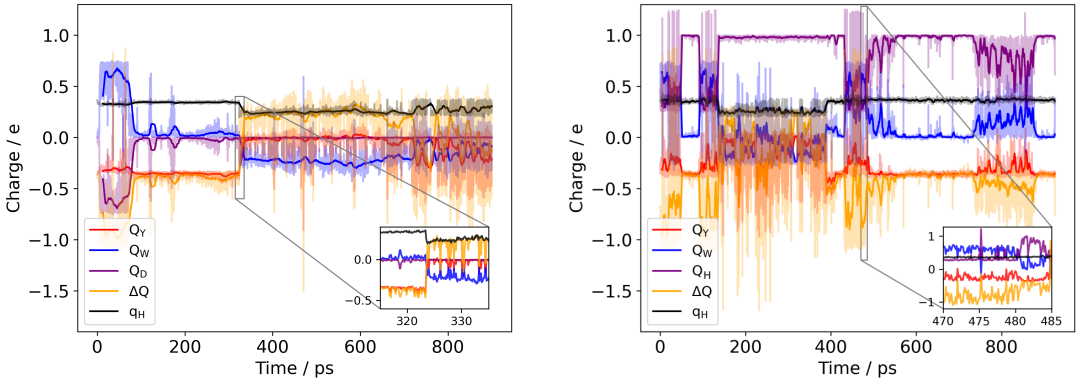
(a) **1a** Proton Transfers

(b) **1b** Proton Transfers

Figure 4.6: FES of two-dimensional proton transfer in MmCPDII: (a) reaction **1a** and (b) reaction **1b**. Inlets show representative structures of the reactant (R), intermediate (I), and product (P) states.

an unstable intermediate "I". The subsequent proton transfer from Y345 to W388$^{\bullet}$ overcomes a high reaction free energy barrier and leads to a product state "P" that is lower in energy than the intermediate but still significantly higher than the reactant. In comparison, in reaction **1b**, both proton transfer steps exhibit similar reaction free energy barriers of $\leq 10\,\mathrm{kcal\,mol^{-1}}$. Both the reactant and intermediate states are at similar energy levels, but the final product is the lowest.

In Fig. 4.7, the partial charge analysis for reactions **1a** and **1b** is shown. In the reactant state of reaction **1a** aspartate is negatively charged while tryptophan is positively charged. W388$^{\bullet+}$ and D321^{-} have partial charges of $\pm 0.7\,\mathrm{e}$, as the hydrogen ($q_{H2} \sim 0.3\,\mathrm{e}$), which is transferred in PT1 is assigned to D321^{-} due to technical reasons. After PT from W388$^{\bullet+}$ to D321^{-}, both residues reach a neutral charge (0 e), while $Q_Y = -0.4\,\mathrm{e}$. Consequently, ΔQ increases from $-1\,\mathrm{e}$ in the reactant state to $-0.4\,\mathrm{e}$ in the intermediate state. When the secondary PT occurs, from Y345 to W388$^{\bullet}$, the partial charges invert, q_H decreases slightly, and ΔQ reaches about 0.3 e in the product state. Fluctuations in ΔQ are attributed to rapid interconversion between Q_Y and Q_W, while Q_D remains stable at 0 e.



(a) 1a Partial Charges (W6)

(b) 1b Partial Charges (W1)

Figure 4.7: Analysis of partial charges of residues in the QM region in one particular walker. The insets show magnified views of characteristic transfer regions to better track changes in the partial charges. MmCPDII reactions (a) **1a** and (b) **1b**.

In contrast, in reaction **1b** the charges fluctuate more strongly. In the reactant state, $Q_H = 0.3\,\mathrm{e}$, $Q_W = 0.7\,\mathrm{e}$, and $Q_Y = -0.4\,\mathrm{e}$, which yields $\Delta Q = -1\,\mathrm{e}$. After PT from W388$^{\bullet+}$ to H354, Q_H increases to 1 e, Q_W drops to 0 e, and Q_Y equals $\Delta Q = -0.4\,\mathrm{e}$ in the intermediate state. Following the subsequent PT from Y345 to W388$^{\bullet}$, q_H decreases again to 0.3 e, while Q_Y and Q_W interconvert as in reaction **1a**, resulting in a product state with $\Delta Q = 0.3\,\mathrm{e}$.

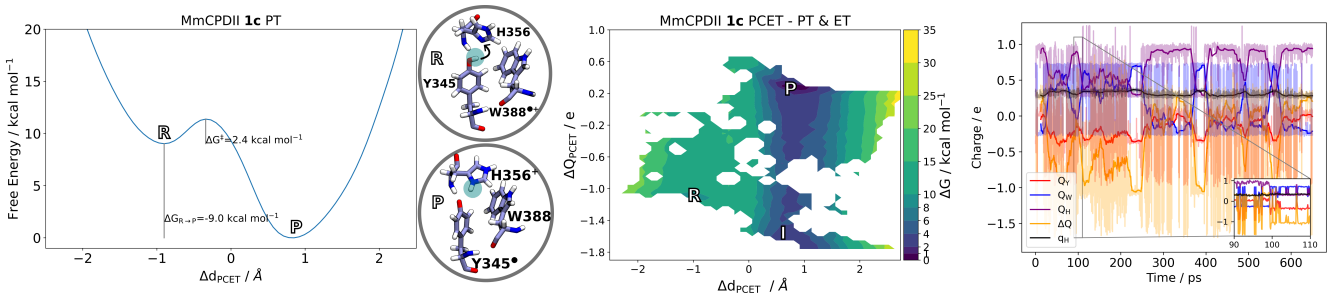
The free energy barrier $\Delta G^{\ddagger}_{\mathrm{PT1}}$ for proton transfer from W388H$^{\bullet+}$ is $5\,\mathrm{kcal\,mol^{-1}}$ lower when the histidine acts as the proton acceptor than when the aspartate does (see $\Delta G^{\ddagger}_{\mathrm{PT1}}$ for reaction **1a** and **1b** in Table 4.2). The intermediate in **1a** is highly

Table 4.2: Free energy barriers for the possible proton transfer reaction in MmCPDII and CraCRY. Values retrieved from WT-MTD FES.

	Reaction	PT1		PT2		(overall) PCET	
		$\Delta G^{\ddagger}_{\mathrm{PT1}}$	$\Delta G_{\mathrm{R}\to\mathrm{I}}$	$\Delta G^{\ddagger}_{\mathrm{PT2}}$	$\Delta G_{\mathrm{I}\to\mathrm{P}}$	$\Delta G^{\ddagger}_{\mathrm{PCET}}$	$\Delta G_{\mathrm{R}\to\mathrm{P}}$
MmCPDII	**1a**	13.5	13.4	8.8	-3.9	22.2	9.5
	1b	9.7	1.1	8.8	-7	15.8	-5.9
	1c					2.4	-9.0
	2a					5.3	-8.3
	2b	2.3	-1.6	9	-5.7	14.7	-7.3
CraCRY	**1a**	22.3	22.3	8.8	-2.7	31.1	19.6
	1b					5.8	5.1
	2a					5.4	-8.8

unstable, lying 14 kcal mol^{-1} above the reactant in free energy. In both reactions, the secondary proton transfer from Y345 to W388$^{\bullet}$ exhibits a reaction free energy barrier of approximately $\Delta G^{\ddagger}_{\mathrm{PT2}} = 9\,\mathrm{kcal\,mol^{-1}}$. Considering the overall reactions, **1a** is both thermodynamically and kinetically unfavorable: the product state lies 10.4 kcal mol^{-1} above the reactant, with an overall reaction free energy barrier of $\Delta G_{\mathrm{R}\to\mathrm{P}} = 23\,\mathrm{kcal\,mol^{-1}}$. In contrast, for reaction **1b** (involving histidine), the product state is more stable than the reactant ($\Delta G_{\mathrm{R}\to\mathrm{P}} = -5.9\,\mathrm{kcal\,mol^{-1}}$), with a reaction free energy barrier of $\Delta G^{\ddagger} = 15.8\,\mathrm{kcal\,mol^{-1}}$.

We also considered the possibility of a PCET mechanism involving distinct proton and electron acceptors. In this case, Y345 donates its proton to H354 and its electron to W388H$^{\bullet+}$ (see reaction mechanism **1c** in Fig. 4.5). Fig. 4.8(a) indicates that this reaction is energetically favorable, with a low free energy barrier of $\Delta G^{\ddagger}_{\mathrm{PCET}} = 2.2\,\mathrm{kcal\,mol^{-1}}$ and a product state $\Delta G_{\mathrm{1c:R}\to\mathrm{P}} = -9.4\,\mathrm{kcal\,mol^{-1}}$ lower in energy than the reactant. The FES featuring the electron transfer CV (Fig. 4.8(b)) indicates a minimum around $-1.05\,\mathrm{e}$ corresponding to the reactant state, as $Q_Y = -0.35\,\mathrm{e}$ and $Q_W = 0.7\,\mathrm{e}$, yielding $\Delta Q = Q_Y - Q_W = -1.05\,\mathrm{e}$. Fig. 4.8(c) shows that overall the charges are fluctuating constantly, resulting in the wide basins on the FES. The product state corresponds to the minimum at 0.3 e, in which $Q_Y = 0\,\mathrm{e}$ (radical tyrosine), $Q_W = -0.2\,\mathrm{e}$ and $Q_H = 1\,\mathrm{e}$ (doubly protonated histidine). Another distinct minimum is at $-1.7\,\mathrm{e}$, representing the Y345^{-}/W388^{+} state, corresponding to the intermediate of a stepwise PCET mechanism. The FES demonstrates that multiple PCET mechanisms are possible: either PCET occurs in a concerted fashion, or in a stepwise manner.



(a) 1c Proton Transfer **(b) 1c** PT & ET **(c) 1c** Partial Charges (W4)

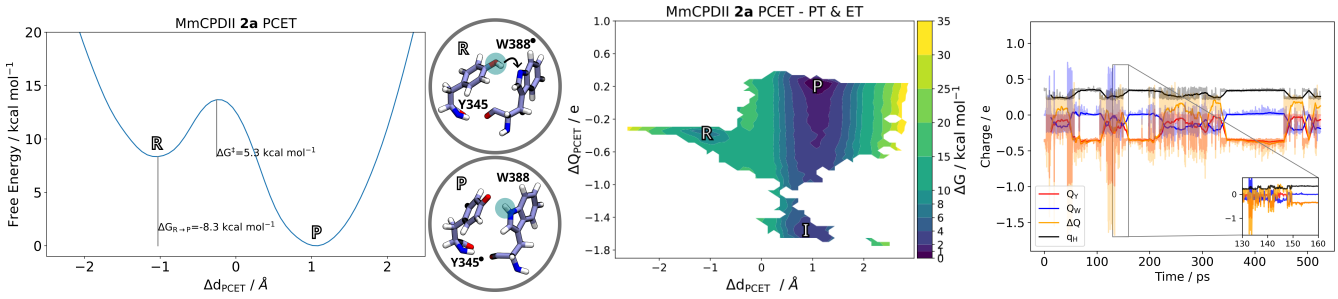
Figure 4.8: Free energy surfaces and partial charge analysis of reaction **1c** of MmCPDII. (a) FES of proton transfer with inlets showing representative structures of the reactant (R) and product (P) states. (b) 2D-FES with PT CV on the x-axis and recalculated ET CV on the y-axis showing two possible PCET mechanisms, either concerted PCET following the minimum-energy pathway (MEP) from the reactant to product minimum, or stepwise PCET through an intermediate state (I) corresponding to Y345^{-}/W388^{+}. (c) Analysis of partial charges of residues in the QM region in one particular walker, with the inset showing a magnified view of a characteristic transfer region to better track changes in the partial charge.

Our results indicate that H354 is a more viable proton acceptor than D357^{-}, suggesting that reaction **1a** is unlikely to occur in the protein environment. Both reactions **1b** and **1c** appear feasible within the protein; however, the free energy barrier for **1c** is considerably lower. This reaction proceeds via a PCET mechanism involving distinct proton and electron acceptors and offers an overall energy gain of about 3.5 kcal mol^{-1} greater than that of the stepwise proton transfer in **1b**.

4.3.2.2 W388$^{\bullet}$: Neutral Radical Tryptophan in Reactant State

We assumed that W388H$^{\bullet+}$ had deprotonated into bulk water as proposed in Ref. [143], due to the nearby water cluster. In this case, there are two possibilities: direct PCET between W388$^{\bullet}$ and Y345 (reaction **2a**) and protonation of Y345 to H354 followed by proton transfer from H354 to W388 (reaction **2b**). This reaction is based on the assumption that there is an intermediate electron transfer step from deprotonated Y345 to W388$^{\bullet}$.

Fig. 4.9(a) shows the FES of reaction **2a**. We observe that the product state with radical tyrosine instead of radical tryptophan in the reactant state is more favorable ($\Delta G_{\mathrm{2a:R}\to\mathrm{P}} = 8.3\,\mathrm{kcal\,mol^{-1}}$). The free energy barrier of the proton transfer is $\Delta G^{\ddagger}_{\mathrm{PCET}} = 5.3\,\mathrm{kcal\,mol^{-1}}$. As in the PT FES, the 2D FES with both, PT and ET, yields the product state in a deeper basin than the reactant state (Fig. 4.9(b)). The reactant minima is narrow at $\Delta Q = -0.4\,\mathrm{e}$ as W388$^{\bullet}$ is neutral and $Q_Y = -0.4\,\mathrm{e}$ as one can see in Fig. 4.9(c). Proton transfer from Y345 to W388$^{\bullet}$ leaves tyrosine in neutral state and this yields $\Delta Q = -Q_W = 0.3\,\mathrm{e}$ in the product state, which is seen as deep minimum in Fig. 4.9(b). In this state, anticorrelated fluctuations of Q_Y and Q_W,

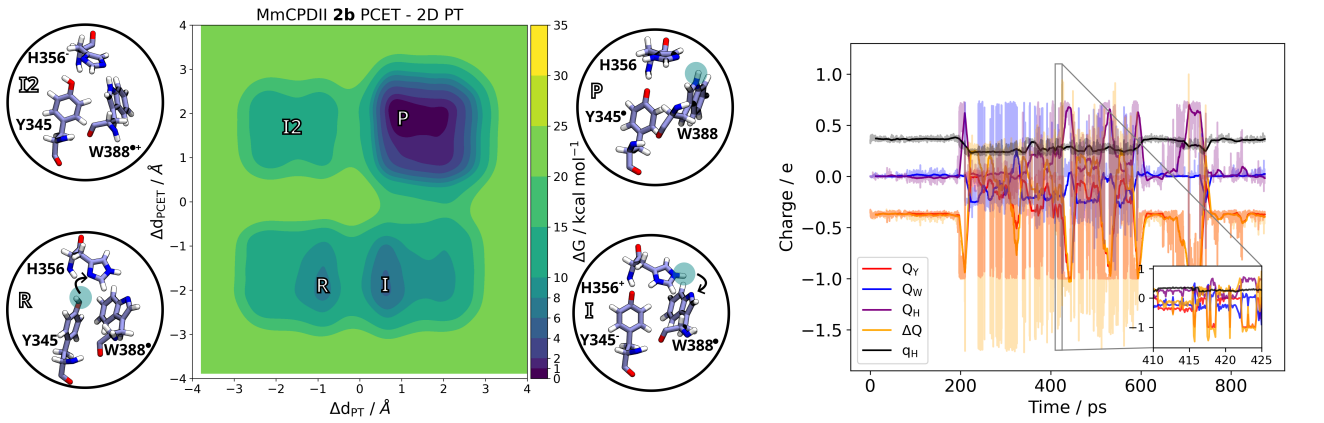


(a) 2a PT **(b) 2a** PT & ET **(c) 2a** Partial Charges (W5)

Figure 4.9: Free energy surfaces and partial charge analysis of reaction **2a** of MmCPDII. (a) FES of proton transfer with inlets showing representative structures of the reactant (R) and product (P) states. (b) 2D-FES with PT CV on the x-axis and recalculated ET CV on the y-axis showing two possible PCET mechanisms, either concerted PCET following the MEP from the reactant to product minimum, or stepwise PCET through an intermediate state (I) corresponding to $Y345^{-}/W388^{+}$. (c) Analysis of partial charges of residues in the QM region in one particular walker, with the inset showing a magnified view of a characteristic transfer region to better track changes in the partial charge.

one decreasing as the other increases, are possible, leading to variations in ΔQ and producing a broader basin stretching from 0.3 to $-0.7\,\mathrm{e}$. Also possible is a state similar to those we observed in PCET between Tyr-Tyr pairs in biomimetic peptides described in our previous publication.[97] There, additional basins at $\Delta Q = \pm 1.3\,\mathrm{e}$ corresponded to intermediate states (Tyr^{+}/Tyr^{-}) as in a step-wise PCET mechanism. Here, our additional basin is at $\Delta Q = -1.7\,\mathrm{e}$ resulting from $Q_Y = -1\,\mathrm{e}$ and $Q_W = 0.7\,\mathrm{e}$, corresponding to the intermediate state ($Y345^{-}/W388^{+}$). The inverted state with $Y345^{+}/W388^{-}$ is not observed along our WT-MTD simulation.

In the case of reaction **2b**, the free energy barrier for the first proton transfer from Y345 to H356 is relatively low, with $\Delta G^{\ddagger} = 3.4\,\mathrm{kcal\,mol^{-1}}$, and the intermediate state lies $2\,\mathrm{kcal\,mol^{-1}}$ below the reactant state (see Table 4.2). The subsequent protonation of W388 by doubly protonated H356 exhibits a free energy barrier of $\Delta G^{\ddagger} = 6.7\,\mathrm{kcal\,mol^{-1}}$, and the product state corresponds to the global minimum of the FES, approximately $8\,\mathrm{kcal\,mol^{-1}}$ lower in energy than the reactant state, as shown in Fig. 4.10(a). An additional, thermodynamically unfavorable state appears in the FES, in which H356 protonates $Y345^{\bullet}$. This reaction is unlikely, as it would result in a negatively charged histidine residue. The state lies about $7\,\mathrm{kcal\,mol^{-1}}$ above the reactant and is only kinetically accessible from the product state. Fig. 4.10(b) shows the charge analysis for this reaction. In the product state, tyrosine is neutral, yielding $\Delta Q = -Q_W = 0.3\,\mathrm{e}$. Several intermediate states are observed. In one, Y345 is negatively charged ($Q_Y = -1\,\mathrm{e}$) while W388 carries a positive charge ($Q_W = 0.7\,\mathrm{e}$), giving $\Delta Q = -1.7\,\mathrm{e}$. In another state, both Q_Y and Q_W are less than $-0.3\,\mathrm{e}$, yielding $\Delta Q \sim 0\,\mathrm{e}$, which corresponds to complete deprotonation of histidine.



(a) 2b Proton Transfers **(b) 2b** Partial Charges (W2)

Figure 4.10: Reactions **2b** of MmCPDII (a) FES with inlets showing representative structures of the reactant (R), intermediate (I), product (P), and second intermediate (I2) states.; (b) Analysis of partial charges of residues in the QM region in one particular walker, with the inset showing a magnified view of a characteristic transfer region to better track changes in the partial charge.

4.3.3 CraCRY

In animal-like cryptochrome from green alga *Chlamydomonas reinhardtii* (CraCRY), Ref. [142] suggests the involvement of $D321^-$ in the PCET mechanism via PT/ET or ET/PT or concerted PCET mechanism (see reaction **1a** and **1b** in Figure 4.11). Another possibility could be the deprotonation of $W322H^{\bullet+}$ to bulk solvent or to an external proton acceptor (such as $D324^-$ or $E374^-$). In the latter case, direct PCET between $W322^\bullet$ and Y373 could occur (**2a**).

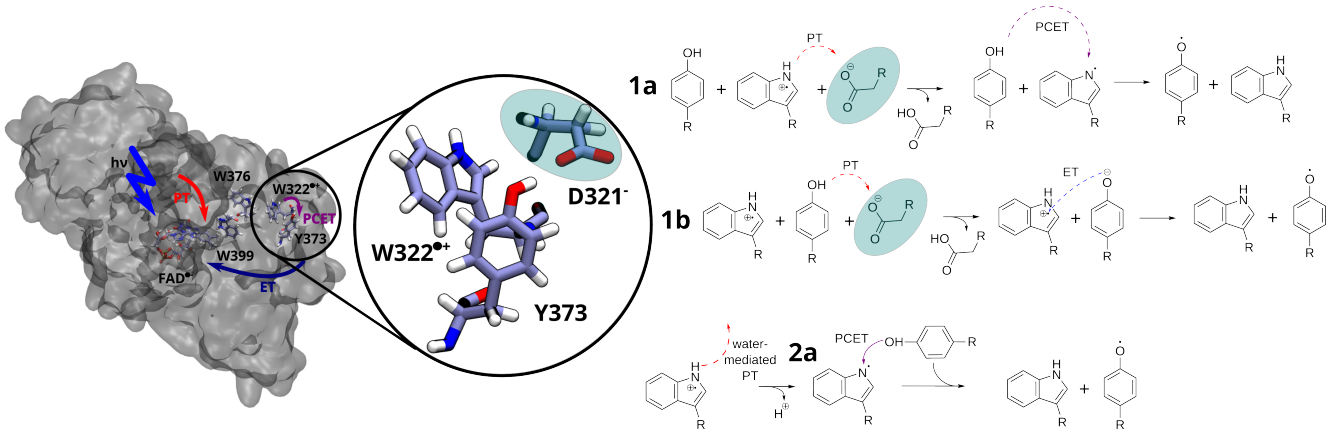
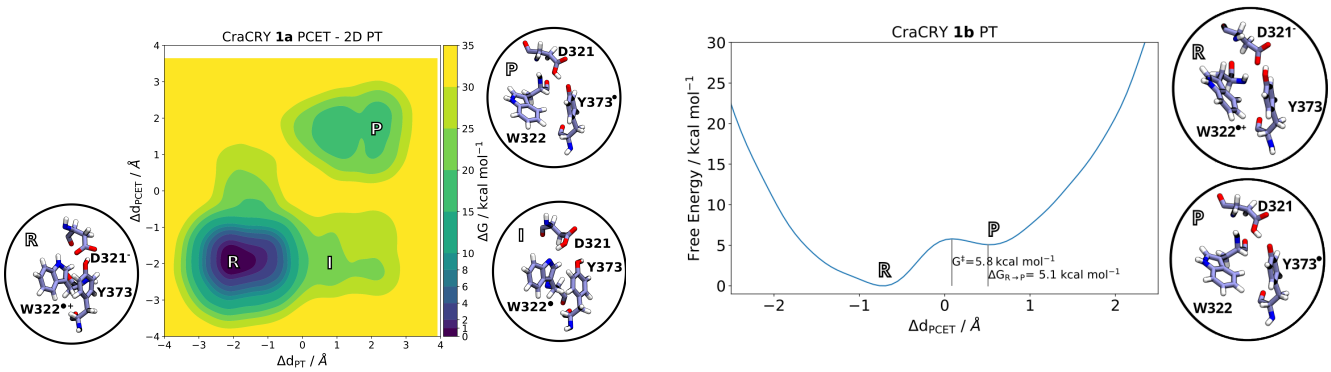


Figure 4.11: Overview of possible PCET reactions in CraCRY: **1a** proton transfer from $W322H^{\bullet+}$ to $D321^-$, followed by PCET between $W322^\bullet$ and Y373; **1b** PCET with different acceptors, where Y373 donates a proton to $D321^-$ and an electron to $W322H^{\bullet+}$; **2a** direct PCET between $W322^\bullet$ and Y373 after prior proton transfer from W322H•+ to an external acceptor.

4.3.3.1 $W322H^{\bullet+}$: Cationic Radical Tryptophan in Reactant State

In the case of CraCRY, we first considered proton transfer from $W322H^{\bullet+}$ to aspartate $D321^-$, followed by proton transfer from Y373 to $W322^\bullet$, see reaction **1a**. Similarly to MmCPDII, proton transfer to aspartate is highly unfavorable: Again, the reactant state is lower in energy than the product (see Fig. 4.12(a)) and the free energy barrier for the overall reaction is $\geq$30 kcal mol^{-1}.



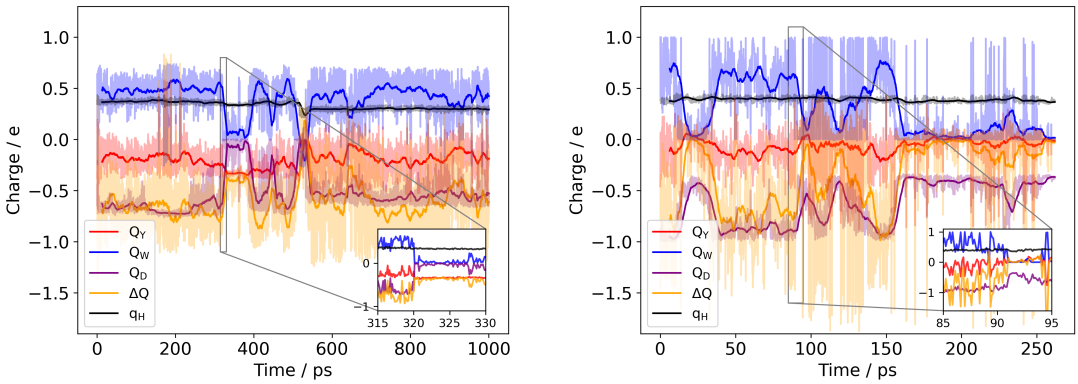
(a) 1a **(b) 1b**

Figure 4.12: FES of PCET reactions in CraCRY with inlets showing representative structures of the reactant (R), intermediate (I), and product (P) states: (a) Two-dimensional proton transfer (reaction **1a**) and (b) Proton transfer (reaction **1b**)

In reaction **1b**, we investigate the role of $D321^-$ as the proton acceptor when Y373 acts as the proton donor. We expect ET from Y373 to $W322^{\bullet+}$ to occur first, followed by deprotonation of the resulting cationic tyrosyl radical, which has an estimated pK_a of -2. Fig. 4.12(b) shows that this reaction is energetically unfavorable: the reactant state lies lower in energy

than the product state, and the product minimum is very shallow, indicating that the product would be unstable. Accordingly, our simulations suggest that no stable product can be formed when the proton is transferred from either W322H$^{\bullet+}$ or Y373 to D321$^-$.

In Fig. 4.13 the charge analysis for reactions **1a** and **1b** is shown. In reaction **1a**, in the reactant state the partial charges of W322H$^{\bullet+}$ and D321$^-$ are $\pm 0.7\,e$ as the hydrogen transferred in PT1 is assigned to D321$^-$ for practical reasons. $Q_Y = -0.4\,e$, yielding approximately $\Delta Q = -1.1\,e$. In the intermediate state, W322 and D321 are neutral and $Q_Y = \Delta Q = -0.4\,e$. After the secondary proton transfer from Y373 to W322, tyrosine is neutral, yielding $\Delta Q = -Q_W = 0.7\,e$ in the product state. In the case of reaction **1b**, W322H$^{\bullet+}$ and D321$^-$ are $\pm 1\,e$ in the reactant state. $Q_Y = -0.4\,e$, which yields $\Delta Q = -1.4\,e$. The charges of Q_W and Q_Y are fluctuating strongly, yielding fluctuation of ΔQ between -1.4 and 0.3 e. Proton transfer between Y373 and D321$^-$ results in $Q_D = -0.4\,e$ and tyrosine and tryptophan are neutral, yielding $\Delta Q = 0\,e$ as well.



(a) 1a Partial Charges (W3)

(b) 1b Partial Charges (W2)

Figure 4.13: Analysis of partial charges of residues in the QM region in one particular walker. The insets show magnified views of characteristic transfer regions to better track changes in the partial charges. CraCRY reactions (a) **1a** and (b) **1b**.

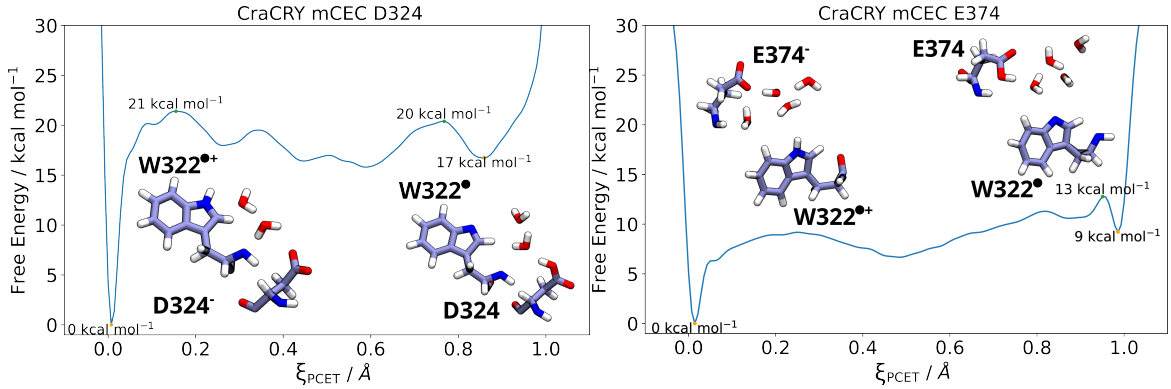
4.3.3.2 Deprotonation of W388H$^{\bullet+}$ to External Proton Acceptor via a Water Chain

Since no nearby histidine residues are available to serve as proton acceptors in the PCET reaction, we considered PT from W322H$^{\bullet+}$ to more distant, external acceptors. In this case, PT proceeds via hydrogen-bonded water molecules, involving multiple bond-breaking and bond-forming events. Such processes require the use of the modified center of excess charge (mCEC) coordinate, which defines the spatial position of the excess proton at any time relative to donor and acceptor groups (see section 2.6.2.1 for details).

The mCEC coordinate was employed to bias PT along a water wire from W322H$^{\bullet+}$ to either D324$^-$ or E374$^-$. Fig. 4.14 shows the resulting FESs for both pathways. For PT to D324$^-$, the reactant state is located in a deep minimum, with the free-energy profile rising steeply to a barrier of about 21 kcal mol^{-1}. The FES exhibits a rugged character, with intermediate fluctuations between 15 to 20 kcal mol^{-1}. Despite occasional transfer events observed in the trajectories, no well-defined product minimum emerges near $\xi = 1$ Å, suggesting that stabilization of the product state with protonated D324 is unlikely.

In contrast, the pathway toward E374$^-$ shows more favorable features. Although the reactant satte remain the global minimum, the free-energy barrier increases only gradually, reaching approximately 13 kcal mol^{-1}. Beyod the barrier, the proton stabilizes in a distinct product basin, with the product state lying about 9 kcal mol^{-1} higher in free energy than the reactant.

Taken together, these results indicate that E374$^-$ is more likely proton acceptor, given its lower barrier and stabilization of the product state, in contrast to D324$^-$ where no product minimum is observed. Nevertheless, it is important to note that neither residue must represent the final proton sink. Further deprotonation into bulk water, though technically unfeasible to simulate, remains a plausible scenario in the protein environment.



(a) W322$^{\bullet+}$/D324^{-} **(b)** W322$^{\bullet+}$/E374^{-}

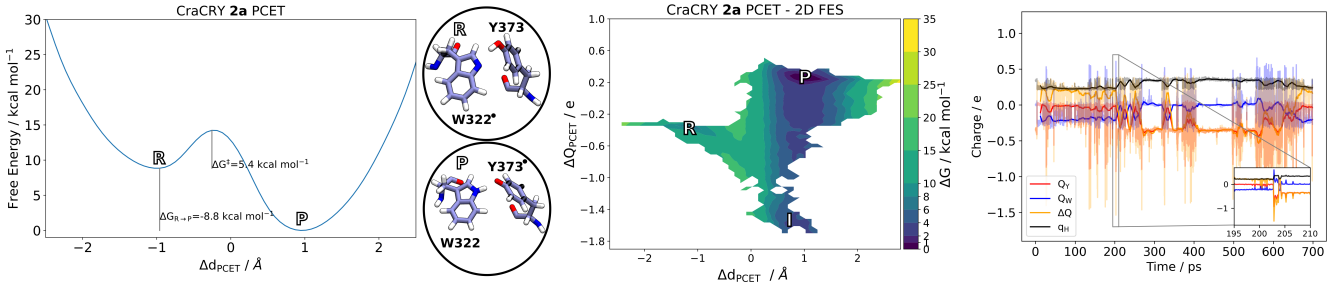
Figure 4.14: 1D FESs of W322H$^{\bullet+}$ deprotonation to external proton acceptor: (a) D321^{-} and (b) E374^{-} based on the mCEC coordinate along a water wire. Representative structures of the initial and final states are shown.

4.3.3.3 W322$^{\bullet}$: Neutral Radical Tryptophan in Reactant State

Although reaction **1b** exhibits a relative low reaction barrier $\Delta G^{\ddagger} = 5.8\,\mathrm{kcal\,mol^{-1}}$, the absence of a stable product state makes this pathway unlikely. Instead, it appears more consistent that W322H^{+} donates its proton to E374^{-}, and subsequently interacts as the neutral radical species W322$^{\bullet}$ with tyrosine Y373 (reaction **2a**).

In this reaction, the free energy barrier is quite low with $\Delta G^{\ddagger} = 5.4\,\mathrm{kcal\,mol^{-1}}$ and the product state lies $\Delta G_{\mathrm{R}\to\mathrm{P}} = -8.8\,\mathrm{kcal\,mol^{-1}}$ lower in energy than the reactant state (see Fig. 4.15(a)). In Fig. 4.15(b) the 2D FES is shown. Here, a narrow minimum at $\Delta Q = -0.4\,\mathrm{e}$ is observed for the reactant state and at $\Delta Q = 0.3\,\mathrm{e}$ for the product state. These minima were already observed in MmCPDII reaction **2a**. Interestingly, the broader basin of the product state extends only from -0.3 to 0.3 e. Additionally, the state corresponding to the stepwise PCET intermediate (Y373^{-}/W322^{+}) is also present. The partial charges analysis yields $Q_W = 0\,\mathrm{e}$ and $\Delta Q = Q_Y = -0.4\,\mathrm{e}$ in the reactant state and $Q_Y = 0\,\mathrm{e}$ and $\Delta Q = -Q_W = 0.4\,\mathrm{e}$ in the product state (see Fig. 4.15(c)) In intermediate state of the stepwise reaction $Q_W = 0.6\,\mathrm{e}$ and $Q_Y = -0.9\,\mathrm{e}$, yielding $\Delta Q = -1.5\,\mathrm{e}$.

Overall, we conclude that deprotonation to the bulk solvent or an external proton acceptor is more likely in CraCRY than direct involvement of D321 in the PCET mechanism, due to high reaction barriers and the instability of the product states.



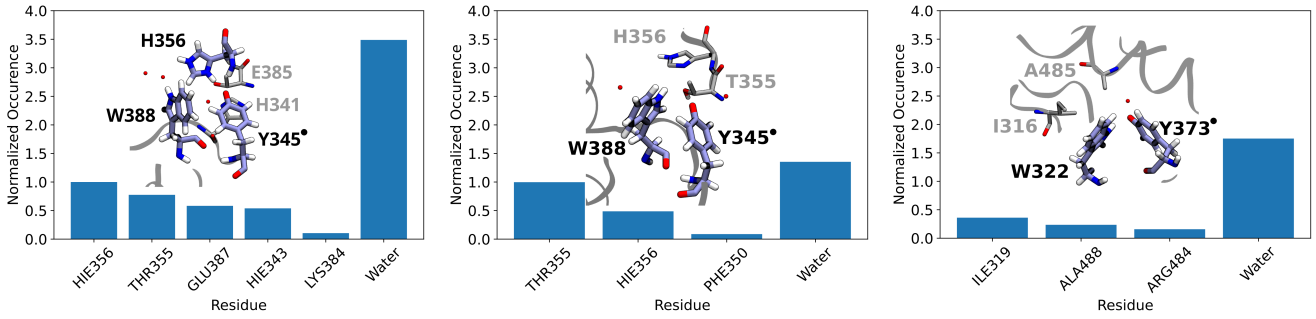
(a) 2a PT **(b)** 2a PT & ET **(c)** 2a Partial Charges (W5)

Figure 4.15: Free energy surfaces and partial charge analysis of reaction **2a** of CraCRY. (a) FES of proton transfer with inlets showing representative structures of the reactant (R) and product (P) states. (b) 2D-FES with PT CV on the x-axis and recalculated ET CV on the y-axis showing two possible PCET mechanisms, either concerted PCET following the MEP from the reactant to product minimum, or stepwise PCET through an intermediate state (I) corresponding to Y373^{-}/W322^{+}. (c) Analysis of partial charges of residues in the QM region in one particular walker, with the inset showing a magnified view of a characteristic transfer region to better track changes in the partial charge.

4.3.4 Structural Comparison

A more detailed structural analysis was performed for the reactions of actual interest in the proteins, due to the lower free energy barriers: reactions **1c** and **2a** in MmCPDII and reaction **2a** in CraCRY.

Figure 4.16 shows the occurrence of residues in proximity to the reaction center during the metadynamics simulations of reactions **1c** and **2a** in MmDPCII, and reaction **2a** in CraCRY. Table 4.3 reports the average number of hydrogen bonds formed with residues participating in the PCET reaction.



(a) MmCPDII **1c** **(b)** MmCPDII **2a** **(c)** CraCRY **2a**

Figure 4.16: Normalized occurrence of amino acids and water molecules within 3 Å of the atoms involved in the PCET mechanism for reactions **1c** and **2a** in MmDPCII, and reaction **2a** in CraCRY. Only residues with an occurrence ≥ 0.05 are displayed.

In reaction **1c** of MmCPDII, the number of hydrogen bonds to water is comparable to that in reaction **2a** in CraCRY, despite a higher number of water molecules being present in the vicinity. By contrast, in MmCPDII reaction **2a**, the number of hydrogen bonds to water is reduced by nearly half. Although H356 is not explicitly included in the QM region in reaction **2a**, it still resides close to W388 and Y345. This indicates that H356 naturally remains nearby, even without the artificial distance restraints applied in reaction **1c** to facilitate the proton transfer. However, in reaction **2a**, T355 plays a crucial role in stabilizing the radical tyrosine: following proton transfer from Y345 to W388$^\bullet$, the distance between T355 and Y345 decreases, enabling the formation of a stabilizing hydrogen bond.

There are many more water molecules in the vicinity of the reaction center in reaction **1c** of MmCPDII than in the other two reactions. This could be due to the additional proton transferred to H356 when selecting atoms for the analysis of neighboring residues and nearby waters.

In both reactions, **1c** and **2a**, approximately 80% of the hydrogen bonds to amino acids involve a single residue: H356 in **1c** and T355 in **2a**. By contrast, in CraCRY reaction **2a**, no hydrogen bonds are formed with other amino acids apart from the key interaction between W322 and Y373 or with water molecules. In CraCRY, the residues involved in the PCET are more closely positioned and less influenced by neighboring amino acids, as the local environment is less crowded compared

Table 4.3: Average number of hydrogen bonds of atoms involved in the PCET reaction either to water molecules (solvent), to other amino acids residues (protein), or between the partners in the PCET reaction (Trp$^\bullet$-Tyr). Criteria to be met for hydrogen bonding: (i) $\text{angle}_{\text{H-Donor}\cdots\text{Acceptor}} \leq 30°$ and (ii) $\text{distance}_{\text{Donor}\cdots\text{Acceptor}} \leq 3.5$ Å. X=H354 in MmCPDII **1c** and X=T355 in MmCPDII **2a**.

	MmCPDII **1c**	MmCPDII **2a**	CraCRY **2a**
Solvent	1.05	0.57	0.96
Protein	1.31	0.84	0.00
Tyr-Trp$^\bullet$	0.00	0.34	0.51
Tyr-X	1.00	0.73	-
total	2.36	1.75	1.47

to MmCPDII. This reduced number of stabilizing interactions may explain why the phase space explored in the FES of CraCRY reaction **2a**, particularly in the transition region, is less broad than in MmCPDII reaction **2a**.

As expected, the distance between Tyr and Trp is larger in MmCPDII reaction **1c**, since the proton is transferred to His, not Trp. Additionally, there is no hydrogen bond between W388 and Y345 in this case. However, this does not reduce the efficiency of the PCET, as this reaction exhibits the lowest free energy barrier among all systems studied. In contrast, for the two **2a** reactions, the distances between the involved residues are very similar, in line with the comparable free energy barriers observed for these reactions. Fig. 4.17 shows the distances between the centers of mass of the residues involved in the PCET reactions.

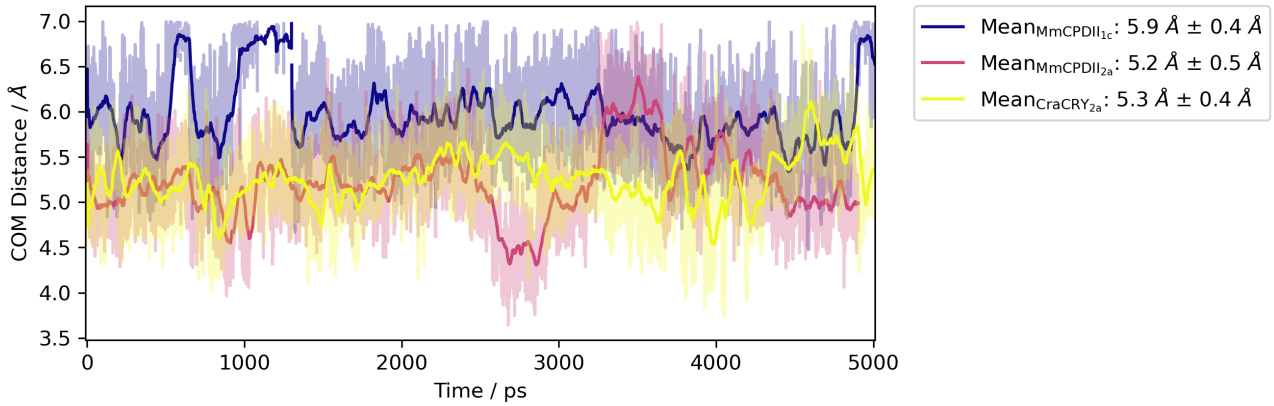


Figure 4.17: Distance between center of mass (COM) of aromatic side chains of residues involved in PCET.

4.4 Conclusion

We applied our PCET simulation protocol from chapter 3 to proteins that are in the spotlight of recent research interest, enabling detailed characterization of PCET reactions involving multiple reaction partners. The use of a semi-empirical QM method allows for efficient treatment of large QM regions. This is essential for capturing multiple PCET partners and potential water-mediated pathways, while coupling with metadynamics ensures thorough sampling of the conformational landscape. This demonstrates the strength of our approach as a powerful and efficient alternative to existing computational strategies for investigating PCET mechanisms in complex protein environments.[196–198]

Experimental studies report the formation of tyrosyl radicals in higher abundance than tryptophanyl radicals in MmCPDII.[143] We can confirm this observation with our computational approach. The product state with radical tyrosine is energetically more favorable than the reactant with radical tryptophan, which we also observed previously for β-W14 in chapter 3. In MmCPDII, two nearby amino acids could accept the tryptophan proton: histidine and aspartate. The pK_a of the histidine side chain is 6.0, whereas that of aspartate is 3.9. If ΔG is estimated from these values, the difference in free energy should be ΔG=2.9 kcal mol^{-1}, which is slightly lower than our observed $\Delta\Delta G^{\ddagger}_{\mathrm{PT1}} = 3.8\,\mathrm{kcal\,mol^{-1}}$ when comparing the deprotonation step of reactions **1a** and **1b**. In both proteins we observe that protonation of aspartate is unfavorable. In MmCPDII, participation of a histidine residue is required (reaction **1c**), whereas in CraCRY, deprotonation of $W322H^+$ must occur first (reaction **2a**) to enable the PCET process.

Due to the choice of our QM method, the free energy barriers may be underestimated because the MD simulations are affected by overdelocalization of the electron density. This is a well known consequence of the self interaction error inherent to DFTB3, which relies on GGA-type functionals.[61,63,147–149] However, since we compare different mechanisms and barriers within the same system, our conclusions are based on relative rather than absolute values. As a result, these systematic errors are expected to cancel out and should not affect the outcome of our study.

A limitation of our current approach is that only PT was explicitly biased during metadynamics, while the ET coordinate remained unbiased. As a result, the FESs involving the electron transfer CV may not be converged. This limits the ability to draw quantitative conclusions about the electron transfer mechanism. In particular, for systems where more than two residues participate in the PCET mechanism, the FESs involving both proton and electron transfer were not yet fully converged. A promising solution would be to apply a bias potential directly on the ET coordinate, as enabled by the CP-DFTB method.[60,66] The following chapter 5 discusses this approach in detail and addresses how to overcome its associated methodological challenges.

5 Biasing the Electron Transfer: Coupled Perturbed Equations in DFTB

5.1 Introduction

As outlined in the previous chapters, understanding proton-coupled electron transfer (PCET) reactions in biological systems requires efficient sampling of both proton and electron transfer coordinates. Metadynamics (MTD), when combined with semi-empirical methods such as density functional tight binding (DFTB), offers a powerful approach for this purpose. Whereas biasing the proton transfer (PT) collective variable (CV) is straightforward and there are different approaches, even if a water-chain is involved,[81] biasing the electron transfer (ET) CV remains a challenge. A major limitation of the protocol employed in the previous two chapters is its inability to explicitly enhance sampling of the ET CV. Alternative strategies, such as constrained density functional theory (DFT),[150] have been used to address this issue. More recently, the implementation of coupled-perturbed (CP) equations[64] within the DFTB framework has enabled direct biasing of ET CVs in MTD simulations,[65,66] offering a promising route to overcome this limitation. This approach allows for a direct bias on the electronic structure rather than relying only on bias on the proton transfer. However, limitations were revealed: when CP equations are solved only for the quantum mechanical (QM) region, significant artifacts appear, such as artificially increased barrier heights.[66] This issue was mitigated by including molecular mechanics (MM) atoms in the CP treatment within a hybrid quantum mechanics/molecular mechanics (QM/MM) scheme, which comes at considerable computational cost, however.[60]

This chapter explores strategies to make CP-DFTB-based ET CVs feasible for larger, biologically relevant systems. Specifically, we aim to find out whether limiting the MM region included in the CP calculations via a cutoff can preserve accuracy while reducing computational cost.

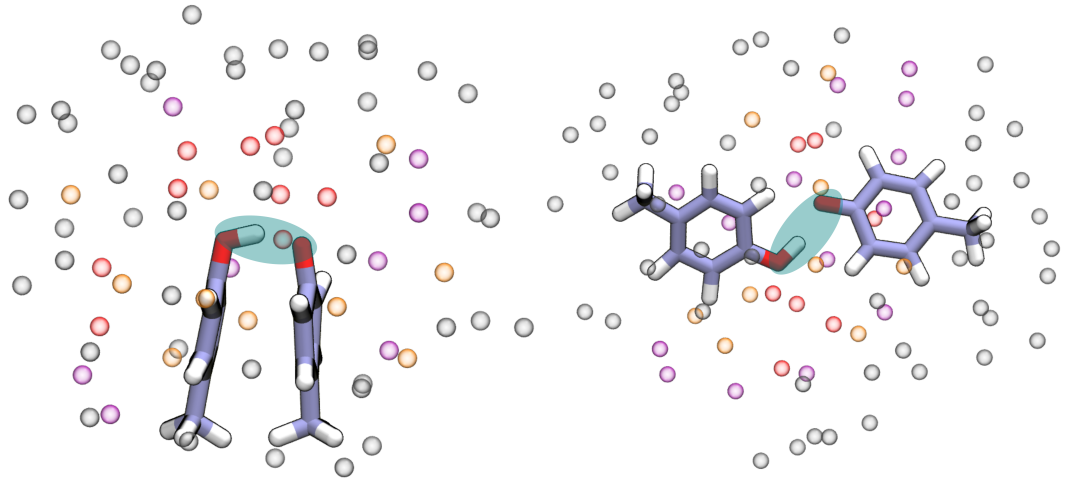
Additionally, we highlight new challenges that emerged in applying this methodology to biomimetic PCET models, including issues of convergence and artificial barriers in non-π-stacked geometries. After testing various cutoff schemes on established model systems, we demonstrate the limitations of the current approach and return to simple test cases (e.g., OH radical in water) to further clarify the source of these issues. Ultimately, this work aims to guide future methodological improvements for reliable biasing of the ET CV in QM/MM metadynamics.

5.2 Computational Implementation

5.2.1 System Preparation and QM/MM Setup

5.2.1.1 Tyrosine Dimer

The tyrosine dimer systems consist of two tyrosine side chains, one of which is in a deprotonated radical state. The structures and parameter files were adapted from previous studies on PCET reactions in Tyr-Tyr$^\bullet$-systems.[60] This model mimics components of the electron transfer chain in ribonucleic reductase (RNR), in which several PCET steps occur.[17] Specifically, the stacked conformation (Fig. 5.1(a)) refers to the configuration αY731 and αY30, whereas the flipped conformation (Fig. 5.1(b)) refers to βY356 and αY30.[98,199] The dimer was enclosed in a sphere of water molecules to improve solvation around the reaction center, which was larger compared to earlier work. The center of the water sphere was moved from the center of mass of the Tyr-Tyr$^\bullet$ complex to the atoms directly involved in the PCET reaction. The simulation box was cubic and significantly larger than the solute system, ensuring that periodic boundary conditions did not affect the dynamics of the water sphere or the tyrosine residues. Also, non-bonded interactions were treated with a cutoff.



(a) Stacked tyrosine dimer **(b)** Flipped tyrosine dimer

Figure 5.1: Representation of the two tyrosine dimers setups containing two tyrosine molecules in water. Color indicates the cumulative set of water side chains for which the CP-equations were solved: red - within a 4 Å bubble; orange - additional molecules included in the 5 Å bubble; purple - additional molecules included in the 6 Å bubble; gray - all remaining water molecules in the "all" setup.

A stochastic dynamics integrator was employed with a time step of 0.5 fs. To maintain solvation while allowing flexibility near the reaction center, spherical restraints were applied to oxygen atoms of water molecules located between 6 and 8 Å from the atoms involved in the PCET reaction, using a force constant of 100 000 kJ mol^{-1} nm^{-2}. This approach was chosen to restrict the outer solvent shell while allowing water molecules in the first solvation shell to remain flexible and adaptable to the PCET process, conditions that would be suppressed if all water molecules were fully restrained. To preserve the geometry of the reaction center, two upper wall restraints at 3 Å were applied on the distance between the oxygen atoms of the tyrosines, and on the sum between the donor–H and acceptor–H distances. Both were enforced using a force constant of 5000 kJ mol^{-1} nm^{-2}.

The two tyrosine molecules were included in the QM region, treated with DFTB3 using the 3ob parameter set,[58,59] while all water molecules were modeled classically using the TIP3P water model.[34] Depending on the simulation setup, the coupled-perturbed equations were solved for all, none, or those water molecules within a cutoff of 4, 5, and 6 Å from the reaction center, in order to investigate computational cost and accuracy.

5.2.1.2 Biomimetic Peptides

The biomimetic peptides were introduced in detail in chapter 3, section 3.2.2 and visualized in Fig. 3.4. In this chapter, we investigate PCET between two tyrosine side chains, applying the CP-DFTB methodology on the β-hairpin Y14 (in both conformations) and the three α-helical proteins. The QM/MM setup and system structures are identical to those described previously.

5.2.1.3 Minimal Setup: $OH^{\bullet}$ in Water

The minimal setup consists of an $OH^{\bullet}$ radical and one water molecule in the QM region (see Fig. 5.2), embedded in a solvent box containing 135 water molecules described with classical force field parameters. To ensure that the solvent remains close

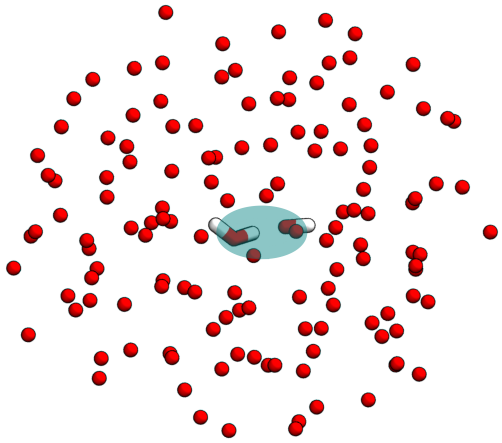


Figure 5.2: Representation of the minimal setup: $OH^{\bullet}$ in water. The CP-equations are solved for all MM water atoms.

to the reaction site, upper-wall distance restraints with a force constant of $1000\,\mathrm{kJ\,mol^{-1}\,nm^{-2}}$ are applied to all water oxygen atoms. The CP-equations are solved for the entire system, including both QM atoms and MM atoms. To systematically explore barrier heights, the distance between the two oxygen atoms in the QM region is restrained to fixed values between 2.6 and 2.8 Å, creating distinct configurations with varying free energy barriers.

5.2.2 Metadynamics Simulations

Two collective variables were used to describe the PCET reaction, as introduced in chapter 3 section 3.2.1: the proton transfer was described using $s_{\mathrm{PT}}(\vec{r}) = \Delta d$, and the electron transfer was captured by $s_{\mathrm{ET}}(\Delta q) = \Delta Q$.

Metadynamics simulations were carried out using a multiple-walker framework, combining GROMACS 2020[125] patched with PLUMED 2.5.1.,[47,117] and DFTB3/3ob[58,59] as implemented in DFTB+,[65] including the CP-DFTB implementation.[64] The initial Gaussian height was set to $0.5\,\mathrm{kJ\,mol^{-1}}$, with widths of $\sigma = 0.02$ Å for the proton transfer CV and $\sigma = 0.02\,\mathrm{e}$ for the electron transfer CV. Both CVs were simultaneously biased during the multiple-walker MTD simulations. Communication between the walkers via the hill files occurred every 500 steps. In the test setup, 24 walkers were employed; for the biomimetic peptides, the number varied depending on the system. Each walker was propagated with a time step of 0.5 fs for a total simulation time of 2.4 ns in the test setup, and at least 4.8 ns in the biomimetic peptides to ensure convergence of the free energy surfaces. Deviations from these general parameters in the biomimetic peptides system are summarized in Table 5.1.

The procedure differed slightly for the minimal setup. Here, we study the influence of the bias on the ET CV along the MTD simulations by simulating three combinations of MTD simulations: (i) biasing only the PT CV, (ii) biasing only the ET CV, and (iii) biasing both, PT and ET CV Calculation of the reweighted bias and the time-dependent factor $c(t)$ was activated by using the keyword CALC_RCT.[44] This provides access to the unbiased probability of the CVs that were not biased during the MTD simulation. Due to the increasing barrier height in the test setup, resulting from the increased distance between $OH^{\bullet}$ and the QM water, the MTD setup must be adapted. For the MTD simulations (i), in which only the proton transfer was biased, one single walker and a total simulation time of 1 ns was sufficient for the convergence of the free energy surface (FES). For MTD simulations (ii), in which only the electron transfer was biased, and MTD simulations (iii), in which both CVs were biased, the MTD parameters were adapted (see Table 5.1).

Table 5.1 gives an overview of the system size, the number of MM atoms for which the CP-equations are solved, and the performance for all systems studied in this chapter.

All data was processed and visualized with python[134] using the matplotlib,[135] NumPy,[136] pandas,[137] and seaborn[138] libraries.

Table 5.1: Overview of the parameters for the WT-Metad simulations and information on the performance of the studied systems.

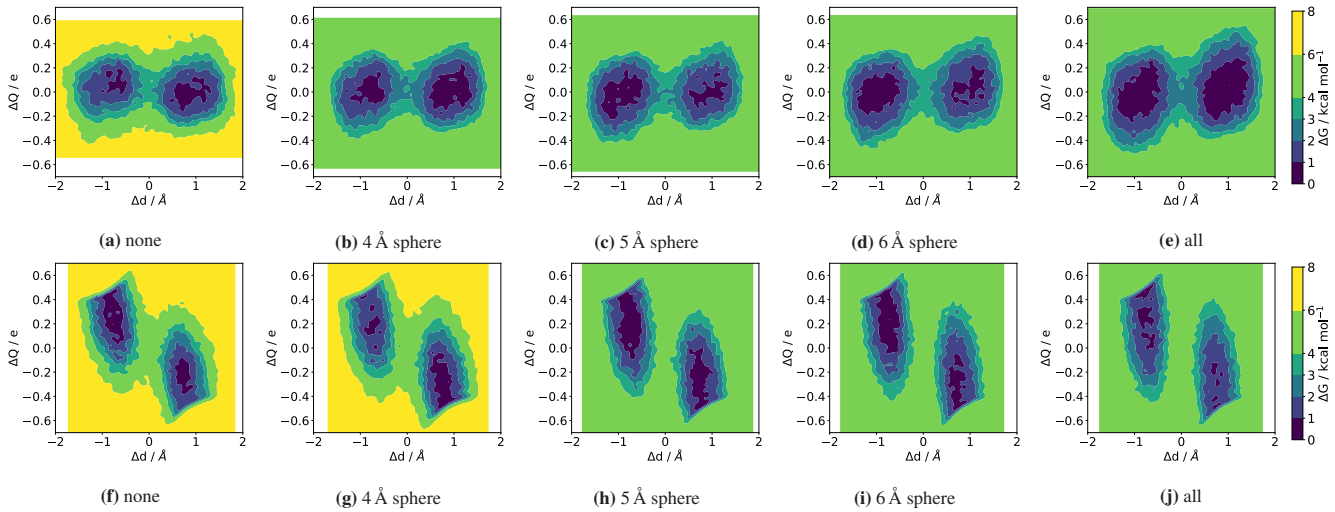
System			WT-Metad parameters					Performance			
			σ_{PT}	σ_{ET}	bf	walker	time	size	AA	water	comp. cost
			nm	e		#	ps	#	#	#	ns/day
Tyrosine dimer	stacked	none	0.02	0.005	-	24	2400	250	-	-	0.019
		4 Å sphere	0.02	0.005	-	24	2400	250	-	24	0.013
		5 Å sphere	0.02	0.005	-	24	2400	250	-	39	0.011
		6 Å sphere	0.02	0.005	-	24	2400	250	-	84	0.007
		all	0.02	0.005	-	24	2400	250	-	219	0.004
	flipped	none	0.02	0.005	-	24	2400	268	-	-	0.018
		4 Å sphere	0.02	0.005	-	24	2400	268	-	18	0.013
		5 Å sphere	0.02	0.005	-	24	2400	268	-	39	0.010
		6 Å sphere	0.02	0.005	-	24	2400	268	-	90	0.006
		all	0.02	0.005	-	24	2400	268	-	237	0.001
β-hairpin peptide	Y14-st	4 Å sphere	0.02	0.005	-	24	4800	11534	18	15	0.010
		5 Å sphere	0.02	0.005	-	48	4800	11534	18	36	0.008
		6 Å sphere	0.02	0.005	-	48	4800	11534	31	87	0.006
	Y14-fl	4 Å sphere	0.02	0.005	-	24	4800	11534	18	18	0.011
		5 Å sphere	0.02	0.005	-	24	4800	11534	31	27	0.010
		6 Å sphere	0.02	0.005	-	48	4800	11534	31	66	0.006
α-helical protein	Y56	4 Å sphere	0.02	0.005	-	48	4800	17644	33	3	0.008
		5 Å sphere	0.02	0.005	-	48	4800	17644	96	12	0.007
		6 Å sphere	0.02	0.005	-	24	5300	17644	78	3	0.006
	Y31	4 Å sphere	0.02	0.005	-	48	5000	17643	25	9	0.009
		5 Å sphere	0.02	0.005	-	24	4800	17643	25	27	0.009
		6 Å sphere	0.02	0.005	-	48	5000	17643	44	12	0.005
	Y30	4 Å sphere	0.02	0.005	-	24	4800	17641	49	-	0.009
		5 Å sphere	0.02	0.005	-	24	4800	17641	58	-	0.008
		6 Å sphere	0.02	0.005	-	48	5300	17641	108	3	0.009
Minimal Setup	(i)	2.60 Å	0.02	-	-	1	100	410	-	405	0.100
		2.65 Å	0.02	-	-	1	100	410	-	405	0.120
		2.70 Å	0.02	-	-	1	100	410	-	405	0.112
		2.75 Å	0.02	-	-	1	100	410	-	405	0.120
		2.80 Å	0.02	-	6	1	400	410	-	405	0.119
	(ii)	2.60 Å	-	0.005	-	1	450	410	-	405	0.085
		2.65 Å	-	0.005	-	1	450	410	-	405	0.123
		2.70 Å	-	0.005	15	8	1000	410	-	405	0.128
		2.75 Å	-	0.005	15	8	1000	410	-	405	0.085
		2.80 Å	-	0.02	21	8	2000	410	-	405	0.085
	(iii)	2.60 Å	0.02	0.02	6	1	400	410	-	405	0.099
		2.65 Å	0.02	0.02	6	1	400	410	-	405	0.119
		2.70 Å	0.02	0.02	6	1	700	410	-	405	0.096
		2.75 Å	0.02	0.02	6	1	1000	410	-	405	0.096
		2.80 Å	0.02	0.02	21	8	500	410	-	405	0.096

5.3 Results and Discussion

5.3.1 Tyrosine Dimer: Influence of Cutoff on PCET Landscape

The computational demand of solving the CP-equations for all MM atoms becomes prohibitive when the total system size (QM and MM region combined) exceeds approximately 350 atoms. To reduce computing time, we introduced a cutoff sphere around the QM region: within this sphere, the CP-equations are explicitly solved for MM atoms, while atoms outside the cutoff influence the PCET reaction only via electrostatic embedding. This cutoff approach was systematically tested to evaluate its impact on the topology of the FES and the energy barriers.

Fig. 5.3 shows the FESs of the tyrosine dimers[60] for the different cutoff variants and the two geometries, the upper panels show the stacked geometry, the lower panels the flipped geometry. The free energy barriers are displayed in Table 5.2.



(a) none **(b)** 4 Å sphere **(c)** 5 Å sphere **(d)** 6 Å sphere **(e)** all

(f) none **(g)** 4 Å sphere **(h)** 5 Å sphere **(i)** 6 Å sphere **(j)** all

Figure 5.3: 2D FES of the PCET in tyrosine dimers for different cutoff spheres of MM atoms for which the CP-equations are solved. Upper panels: stacked tyrosine dimer with reaction barrier height $\Delta G^{\ddagger} \approx 3.0\,\mathrm{kcal\,mol^{-1}}$; lower panel: flipped tyrosine dimer with reaction barrier height $\Delta G^{\ddagger} \approx 4.5\,\mathrm{kcal\,mol^{-1}}$. No notable influence is observed on the landscape of the PCET reaction if the cutoff sphere is enlarged.

For the stacked conformation, the free energy barriers are approximately $3.5\,\mathrm{kcal\,mol^{-1}}$, while they are slightly higher for the flipped conformation, around $4.5\,\mathrm{kcal\,mol^{-1}}$. The variations between different cutoff conditions remain below $1\,\mathrm{kcal\,mol^{-1}}$, which falls within the acceptable margin of error. No notable influence is observed on the landscape of the PCET reaction if CP-equations are solved for more MM atoms.

Table 5.2: Overview of the free energy barrier height of the tyrosine dimers. Values in $\mathrm{kcal\,mol^{-1}}$. CP-equations are solved for all QM atoms. Size of the cutoff sphere defining the MM atoms for which the CP-equations are solved is indicated in the first column.

MM atoms (CP-eq)	$\Delta G^{\ddagger}$ / $\mathrm{kcal\,mol^{-1}}$ Test Setup	
	stacked	flipped
none	3.3	4.4
4 Å sphere	2.6	4.5
5 Å sphere	2.7	4.3
6 Å sphere	3.0	4.5
all	2.4	4.8

The topologies of the free energy landscapes closely resemble those reported in Ref. [60] for an oxygen-oxygen distance of 2.4 Å. In the stacked conformation, no distinct minima are observed, and the electron transfer coordinate ΔQ spans a broad range in both the reactant and product states, from -0.2 to 0.2 e. In contrast, the flipped conformation shows a ΔQ range of 0.4 to 0 e in the reactant state, and from -0.4 to 0 e in the product state.

We therefore conclude that, at least in the test setup, the number of MM atoms for which the CP-equations are solved does not critically affect the free energy landscape.

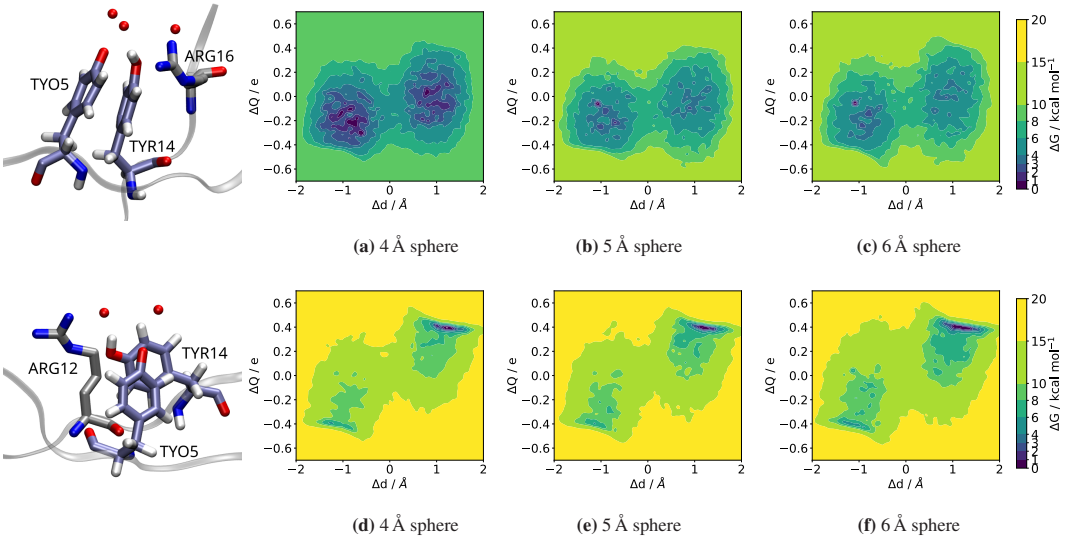
5.3.2 Application to Biomimetic Peptides

The introduction of a cutoff sphere around the atoms involved in the PCET reaction enables the investigation of PCET in larger, biologically relevant systems by significantly reducing the computing time. We applied this methodology to the biomimetic peptides, introduced in chapter 3, section 3.2.2. Since the preceding study using the tyrosine dimers revealed no significant differences based on the cutoff sphere size, we conducted the MTD simulations in triplicate using cutoff spheres of 4, 5, or 6 Å around the atoms involved in the PCET process.

The FES for the β-hairpin peptides and α-helical proteins are displayed in Fig. 5.4 and Fig 5.5, respectively. Table 5.3 provides an overview of the corresponding barrier heights $\Delta G^{\ddagger}$. Recall that the free energy barriers for the biomimetic peptides were approximately 4 kcal mol^{-1} for the β-hairpin peptides and ranged between 6 and 8 kcal mol^{-1} for the various α-helical proteins (see Table 3.3 in chapter 3 section 3.3.3).

Table 5.3: Overview of the free energy barrier height of the biomimetic hairpins. Values in kcal mol^{-1}. CP-equations are solved for all QM atoms. MM atoms for which the CP-equations are solved are indicated in the first column.

	$\Delta G^{\ddagger}$ / kcal mol^{-1}				
MM atoms (CP-eq)	β-hairpin Y14		α-helical proteins		
	Y14-st	Y14-fl	Y56	Y31	Y30
4 Å sphere	4.0	12.5	18.5	12.5	15.5
5 Å sphere	6.5	13.0	15.0	11.3	15.5
6 Å sphere	6.4	12.0	15.0	10.3	17.8



(a) 4 Å sphere **(b)** 5 Å sphere **(c)** 6 Å sphere

(d) 4 Å sphere **(e)** 5 Å sphere **(f)** 6 Å sphere

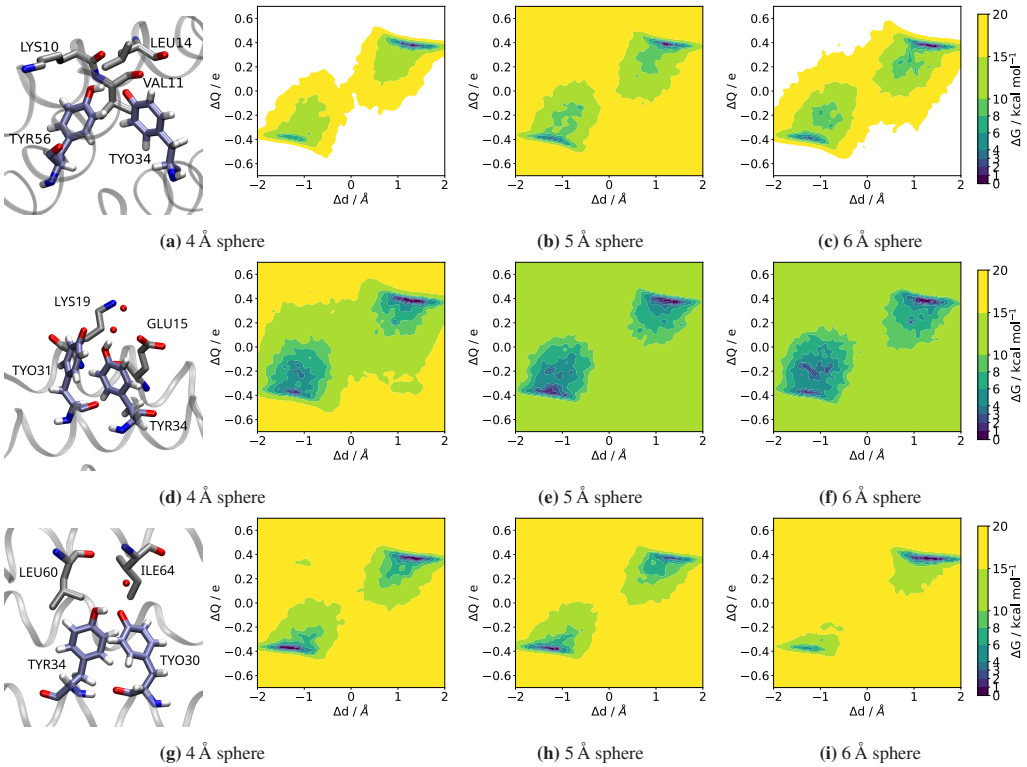
Figure 5.4: 2D FES of the β-hairpin peptides; upper panel: stacked conformation and lower panel: flipped conformation. No notable influence on the landscape of the PCET reaction if CP-equations are solved for MM atoms in a 4, 5 or 6 Å sphere, but large discrepancies for Y14-fl regarding the barrier height as it is increased to $\geq$ 10 kcal mol^{-1}.

The FES of Y14s exhibits a topology similar to that of the stacked conformation investigated in section 5.3.1. Only in the case of the 4 Å cutoff sphere does the free energy barrier match the expected value $\Delta G^{\ddagger} = 4\,\mathrm{kcal\,mol^{-1}}$. Visual inspection of the 2D surface, however, reveals a deep, artificial minimum, likely the result of a suboptimal starting structure. This artifact may account for the observed increase of $\Delta G^{\ddagger}$ by approximately $2\,\mathrm{kcal\,mol^{-1}}$ in the 5 and 6 Å sphere simulations.

By contrast, the β-hairpin Y14f shows an approximately threefold increase in the free energy barrier across all three cutoff spheres. The minima are located at $\Delta Q = -0.6\,\mathrm{e}$ in the reactant state and at $\Delta Q = 0.6\,\mathrm{e}$ in the product state. In the earlier study using reweighted-PT MTD simulations, deep minima were observed at -0.4 and 0.4 e, but only after recalculating the Mulliken charges with long-range corrected DFTB2 (see Fig. 3.9 in chapter 3 section 3.3.2 for details).

For the α-helical proteins, the behavior is similar to Y14f: the free energy barrier doubled for Y56, increased by about $5\,\mathrm{kcal\,mol^{-1}}$ in Y31, and tripled for Y30. Again, there are deep minima at $-0.6\,\mathrm{e}$ in the reactant state and $0.6\,\mathrm{e}$ in the product state. In Y31, the minima are slightly broader than in the other two cases. It is the only α-helical protein, in which the atoms involved in the PCET reaction are solvent exposed, rather than embedded in the protein structure.

Application of the methodology to the biomimetic peptides revealed several limitations. Most notably, the free energy barriers appeared to increase artificially, without any obvious cause. Previous studies have reported that restraining harmonic potentials can introduce systematic errors in the free-energy estimation near the boundary, preventing the system from reaching a stationary state.[200–202] It is therefore possible that the elevated barriers arise from artificial minima induced by the applied upper-wall restraints. Adjusting the restraints and the Gaussian width σ may improve the free energy estimates, an aspect that warrants further investigation. Interestingly, this effect seems to be mitigated in the stacked conformation and, for Y31, by the presence of water molecules in the environment rather than nearby amino acid residues. Additionally, not only are the barriers already artificially high, but the FESs also fail to converge.



(a) 4 Å sphere **(b)** 5 Å sphere **(c)** 6 Å sphere

(d) 4 Å sphere **(e)** 5 Å sphere **(f)** 6 Å sphere

(g) 4 Å sphere **(h)** 5 Å sphere **(i)** 6 Å sphere

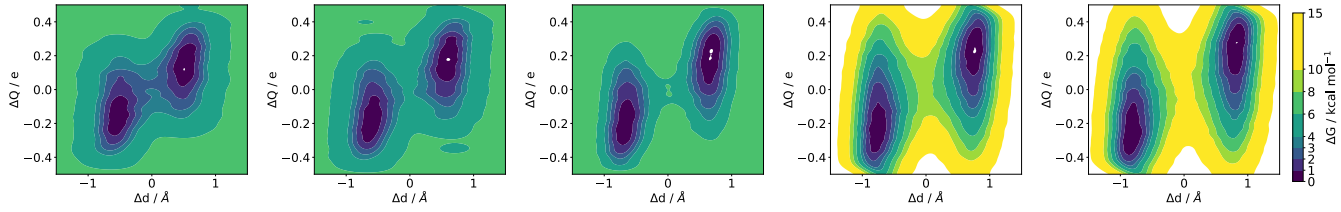
Figure 5.5: 2D FES of the α-helical proteins; upper panel: Y56, middle panel: Y31 and lower panel: Y30. No notable influence on the landscape of the PCET reaction if CP-equations are solved for MM atoms in a 4, 5 or 6 Å sphere, but large discrepancies for all α-helical proteins regarding the barrier height as it is increase to $\geq 10\,\mathrm{kcal\,mol^{-1}}$ in Y31 and even $\geq 15\,\mathrm{kcal\,mol^{-1}}$ in Y56 and Y30.

5.3.3 Minimal Setup: OH• in Water

Considering the methodological challenges outlined above, we returned to a simplified minimal setup to investigate the origin of the artificial barrier increase. In this system, an OH• radical and a single water molecule are treated at the QM level, embedded in a bulk water environment. Owing to the small system size, the CP-equations can be solved for all MM atoms.

To systematically investigate different barrier heights, we performed a series of simulations in which the distance between the oxygen atoms in the QM region was restrained to fixed values between 2.6 and 2.8 Å. For each restrained geometry, MTD simulations were carried out in triplicate under three different biasing schemes: (i) biasing only the proton transfer, (ii) biasing only the electron transfer, and (iii) biasing both proton and electron transfer simultaneously.

Fig. 5.6 displays the FES for the minimal setup under simulation condition (iii), while Table 5.4 lists the corresponding free energy barriers for all three simulation types. As the distance between the two oxygen atoms increases, the barrier height also increases, from $\Delta G^{\ddagger} = 2.5\,\mathrm{kcal\,mol^{-1}}$ at 2.6 Å to $\Delta G^{\ddagger} = 11\,\mathrm{kcal\,mol^{-1}}$ at 2.8 Å. Despite this quantitative change, the overall topology of the FES remains consistent across different oxygen-oxygen distances. The reactant minimum is located at -0.3 to -0.1 e, while the product state ranges from 0.1 to 0.3 e, connected by a narrow minimum energy pathway. Notably, the FES exhibits strict symmetry in all cases.



(a) 2.6 Å **(b)** 2.65 Å **(c)** 2.7 Å **(d)** 2.75 Å **(e)** 2.8 Å

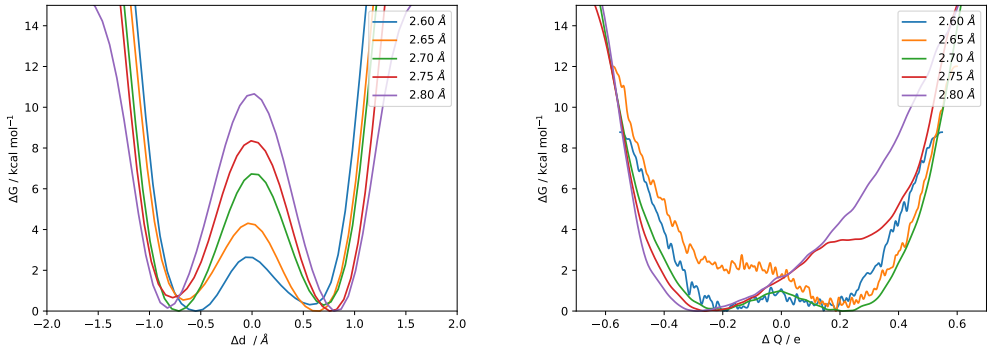
Figure 5.6: 2D FES of the minimal setup with OH• of simulation (iii), in which both PT and ET are biased. Free energy barrier increases with larger distance between the atoms involved in the PCET reaction from (a) $\Delta G^{\ddagger}_{2.6} = 2.7\,\mathrm{kcal\,mol^{-1}}$ to (e) $\Delta G^{\ddagger}_{2.8} = 10.1\,\mathrm{kcal\,mol^{-1}}$.

Fig. 5.7 compares the FES obtained from one-dimensional MTD simulations with the corresponding one-dimensional projections of the two-dimensional FES. Interestingly, in contrast to the results observed for the biomimetic peptides discussed in section 5.3.2, the FES obtained from the 2D-metadynamics simulations in the minimal setup converge to the same barrier heights as those from the 1D simulations.

In these cases, applying distance restraints did not introduce any artificial deep minima in the FES, confirming that the methodology yields accurate results, at least for minimal model systems. By increasing the Gaussian width σ and gradually adjusting the number of walkers, we were able to optimize the computing time relative to the barrier height. Among the three simulation types, simulation (i), which biased only PT, converged rapidly. Simulation (iii), involving two-dimensional MTD, required slightly more time due to the increased complexity but still achieved convergence efficiently. In contrast, simulation (ii), which biased only the ET, exhibited significantly slower convergence. In addition, not only was the convergence slower, but also the simulations crashed before convergence was achieved, as in the case of distance 2.75 Å and 2.80 Å. Notably, all three simulation types yielded the same 2D-FES (after reweighting for simulation types (i) and (ii)). This suggests that directly biasing the ET CV did not offer particular advantages in this system, but instead increased the computational cost without improving convergence. However, convergence using only the PT CV should be seen as a special case, possible here due to the small system size, low barriers, and the affordability of treating the entire system.

Table 5.4: Free energy barrier height $\Delta G^{\ddagger}$ of the OH• simulations. Values in $\mathrm{kcal\,mol^{-1}}$. CP-equations are solved for all atoms in the setup.

2.6			2.65			2.7			2.75			2.8		
(i)	(ii)	(iii)	(i)	(ii)	(iii)	(i)	(ii)	(iii)	(i)	(ii)	(iii)	(i)	(ii)	(iii)
2.6	2.4	2.7	4.5	4.2	4.0	6.2	5.9	5.5	6.4	-	8.1	10.4	-	10.1



(a) (i)

(b) (ii)

(c) (iii) - projection on PT axis

(d) (iii) - projection on ET axis

Figure 5.7: FES of the minimal setup with $OH^{\bullet}$, either only one transfer biased or projected onto one axis. The barriers increase with larger distance between the atom involved in the PCET reaction and the FES obtained from the 2D-metadynamics simulations converge to the same barrier heights as those from the 1D simulations. In the case of (b) simulation (ii) bias on ET only, not all FES are converged.

5.3.4 Computational Cost

Table 5.1 summarizes the simulation performance along with the number of MM atoms included in the CP-equation calculations. As shown in Fig. 5.8, there is a steep decrease in the computational performance with an increasing number of MM atoms. Once the number of MM atoms exceeds approximately 100 MM atoms, assuming a QM region of $\sim$ 30 atoms, the simulations become computationally infeasible.

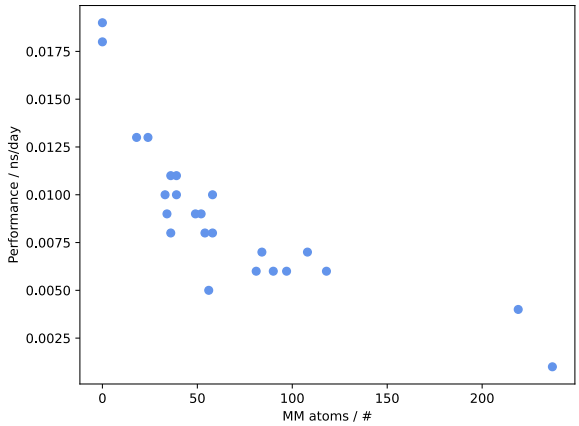


Figure 5.8: Correlation of the number of atoms for which the CP-equations are solved and the computational performance show a steep decrease in computational efficiency with more MM atoms.

A smaller QM region, such as in the minimal setup with only 5 atoms, significantly improves computational performance, allowing simulations with over 400 MM atoms and achieving a tenfold increase in speed. However, 2D MTD simulations (biasing both proton and electron transfer coordinates) require substantially longer timescales than simulations biased only along the proton transfer coordinate. This arises from the increased computational cost associated with biasing multiple CVs, including the expansion of the bias grid and the evaluation of the Gaussians in multiple dimensions. Consequently, for reactions with larger barriers, the number of walkers should be increased, and the Gaussian width (σ) may need to be adjusted to ensure convergence within a reasonable computational time.

5.4 Conclusions

The integration of CP-equations within the DFTB framework offers a major advantage as it allows for explicit enhancement of sampling along the electron transfer CV by applying a bias potential in MTD simulations. This capability provides a more detailed and accurate picture of the PCET process in proteins.

A previous limitation of the method, namely, the need to solve the CP-equations for all MM atoms,[60] which made the simulation of systems larger than approximately 200-400 atoms (depending on the QM region size) unfeasible, was addressed by introducing a cutoff sphere. In this approach, only MM atoms within the defined sphere are included in the solution of CP-equations. Our results show that the number of MM atoms treated in this way does not significantly affect the resulting FES. Consequently, using a smaller cutoff sphere substantially improves computational performance, which decreases steeply with the number of MM atoms included, regardless of the total system size.

In the minimal setup with $OH^{\bullet}$, the 2D MTD simulations converged within a reasonable simulation time and reproduced the barrier heights observed in the corresponding one-dimensional simulations. Moreover, the expected correlation between the increased distance of the oxygen atoms and the free energy barrier was clearly observed. By contrast, simulations of the biomimetic peptides resulted in significantly overestimated free energy barriers, and convergence could not be achieved. Several factors may contribute to this discrepancy: unlike the homogeneous solvent environment in the tyrosine dimers and the minimal setups with $OH^{\bullet}$, the PCET reactions in the peptides occur in a complex and heterogeneous environment composed of amino acid residues and water. Furthermore, the configuration of the residues involved in the PCET can influence the adiabaticity of the reaction, as discussed in Refs. [17, 112, 150], which presents an additional limitation of our approach, as our simulations are restricted to the electronic ground state. Finally, potential effects arising from the applied distance restraints on the PCET-involved atoms and from the MTD parameters have not yet been systematically explored.

Given the radical nature of the PCET systems we investigate, the inclusion of spin-polarization is an important consideration. However, this poses an additional challenge, as spin-polarized CP-equations are not yet implemented within the DFTB framework. Another complication arises in the definition of the electron transfer CV. When the ET CV is described via the Mulliken charge difference, ambiguities can occur, particularly in cases where intermediate states emerge along a stepwise PCET pathway (compare Figs. 3.10 and 4.9(b)). For example, changes in the ET CV are observed even when only H^+ has been transferred. To overcome this, an alternative approach could be to define the ET CV based on spin densities. A related strategy was proposed in Ref. [203], where long-range PT was studied using a combination of a geometrical CV and an electron-density-based CV, the so-called charge transfer factor derived from constrained DFT diabatic states.

An additional drawback of the methodology remains its substantial computational cost. Although the overall system size has only a limited influence on performance, the number of MM atoms included in the CP-equation calculations significantly affects the computational cost, often exceeding practical limits. This constraint becomes especially critical when extending the method to larger proteins such as ribonucleotide reductase, photolyases, or cryptochromes, where PCET processes may involve more than two residues. For instance, in the case of photolyase MmCPDII (see chapter 4 for reference), the complexity of the PCET pathway necessitates an expanded QM region, making the simulations computationally prohibitive. A potential strategy to mitigate this issue, suggested in Ref. [60], is to identify components of the CP-equations that contribute negligibly at longer distances. These components could then be omitted beyond a defined cutoff, reducing the computational burden while maintaining sufficient accuracy.

Therefore, a careful evaluation is necessary to determine when biasing the ET CV is essential. In many systems, one-dimensional MTD simulations that bias only the proton transfer CV may be sufficient, avoiding unnecessary increases in computational cost. However, as demonstrated in the case of MmCPDII in chapter 4, biasing the ET CV is crucial to obtain a converged FES. For such cases, our current protocol appears promising. Nevertheless, further testing is needed, particularly regarding the optimal size of the cutoff sphere, strategies to reduce computational expense, and the suitability of restraints and other simulation parameters.

6 Machine Learning of Absorption Spectra of Retinal Proteins

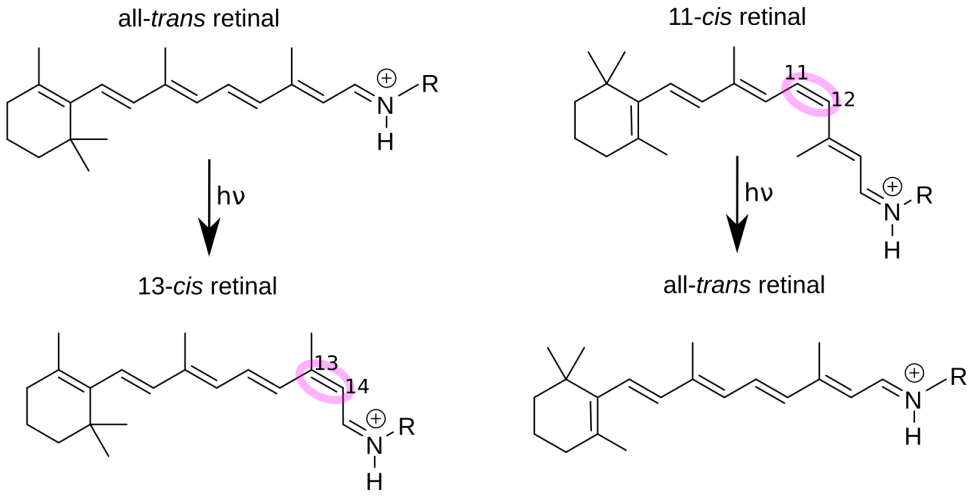
The work presented in chapter 6 was carried out in collaboration with Christian Schmidt and several students.

Author contribution
Christian Schmidt set up the neural network used for training and rerunning the protein structures. Katharina Spies created the parameters for the retinal structures and supervised Béryl Naëmi Greb, Lara Nagel, and Jemina Enkelmann during their bachelor's theses on this topic. Béryl Naëmi Greb prepared the structure and simulations of the rhodopsin protein.[204] Lara Nagel and Jemina Enkelmann prepared a large set of retinal structures in distinct solvent and ionic conditions for 11-cis[205] and all-trans[206] retinal, respectively. Excited-state calculations and structural analysis were performed by Béryl Naëmi Greb during her student assistant position. Katharina Spies prepared the remaining protein structures, trained the machine learning models, and analyzed the neural network results.

6.1 Introduction

Rhodopsins are a family of light-absorbing photoreceptor proteins that play essential roles in vision, bioenergetics, and phototaxis. They enable organisms to sense and respond to light stimuli and have become especially important in the field of optogenetics. In this context, neural activity can be modulated by light-induced de- or hyperpolarization of targeted cells using light-gated ion channels.[207] Applications are diverse, including vision and hearing restoration.[208]

Rhodopsin proteins are classified into two types based on their evolutionary origin: microbial (type I) and animal (type II). The key chromophore in these retinylidene proteins is retinal, which undergoes photoisomerization upon light absorption, from all-trans to 13-cis in type I rhodopsins (Fig. 6.1(a)) and from 11-cis to all-trans in type II rhodopsins (Fig. 6.1(b)).



(a) Type I (microbial) rhodopsin

(b) Type II (animal) rhodopsin

Figure 6.1: Retinal isomerization following photoexcitation: (a) all-trans → 13-cis isomerization in microbial rhodopsin; (b) 11-cis → all-trans isomerization in animal rhodopsin.

This isomerization initiates the photocycle and induces a conformational change in the protein, thereby activating downstream pathways such as ion channel opening.[20] Within the protein, retinal is embedded in an internal binding pocket and covalently attached to the protein via a lysine side chain. In most cases, the retinal Schiff base is protonated in the ground state and stabilized by nearby counterions.[19,209–211]

Across various retinylidene proteins, a broad range of the visible spectrum is covered through a mechanism known as color tuning, reaching e.g. from channelrhodopsin 2 (ChR2), which absorbs at 470 nm (2.62 eV), over bovine rhodopsin at 500 nm (2.48 eV), and bacteriorhodopsin (bR) at 570 nm (2.18 eV) to red-light absorbing Chrimson at 590 nm (2.10 eV). This spectral diversity is enabled by several factors: the planarity of the ring-chain system, electrostatic and dispersion interactions with amino acids in the binding pocket, and, crucially, the interaction between the chromophore and its complex counterion. Fig. 6.2 illustrates the principle of color tuning: Negative charges near the retinal Schiff base stabilize the electronic ground state (S_0), increasing the energy gap to the excited state (S_1) and resulting in blue-shifted absorption. In contrast, negative charges located near the β-ionone ring stabilize the excited state (S_1), lowering its energy and allowing absorption of lower-energy, red-shifted light. The opsin shift is defined as the difference between the maximum absorption wavelength λ_{max} of retinal in vacuo and its λ_{max} when bound to the opsin protein.[19,212]

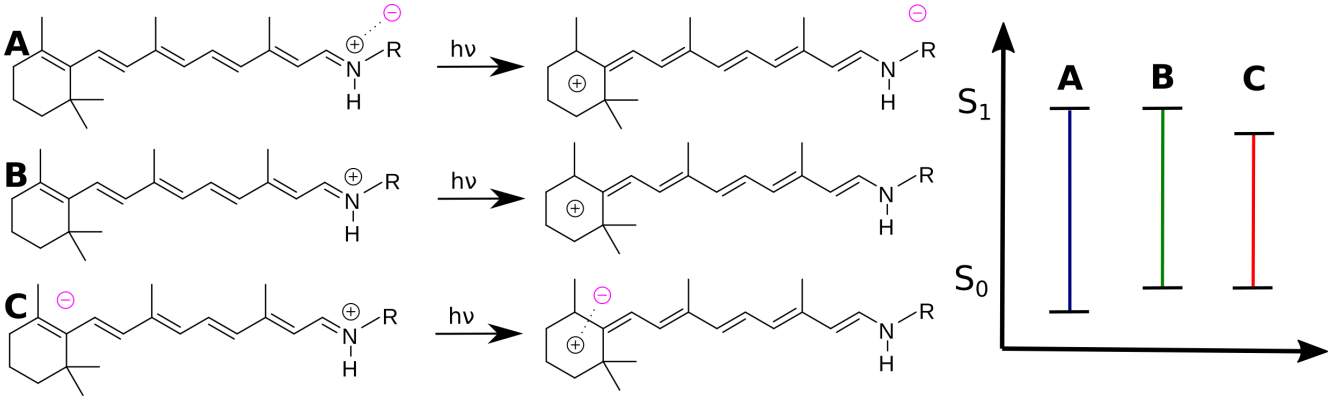


Figure 6.2: Schematic representation of the concept of color tuning (adapted from Ref. [19]). **A** Ground state energy is lowered by stabilization of the protonated retinal Schiff base by a nearby negatively charged counterion. **B** No counterions and therefore no energy level change. **C** Negative counterion near the β-ionone ring stabilized the electronic excited state.

Several theoretical studies have investigated rhodopsins, focusing on the structure and protonation states of titratable amino acids,[213] hydrogen-bonding networks,[21,214–216] photoisomerization,[217] and polarization effects.[218,219] In Ref. [21], a correlation was highlighted between the bond length alternation (BLA), the average length difference between single and double bonds in the conjugated polyene chain, and the excitation energies ω across various rhodopsin proteins.

Hybrid quantum mechanics/molecular mechanics (QM/MM) studies provide detailed insight into the structure and function of photoreactive proteins by accurately modeling the chromophore and its surrounding environment. However, investigating excited-state dynamics remains challenging and computationally demanding.[23]

A promising strategy to overcome these limitations is the use of machine learning-based models. Section 2.5 introduces the application of machine learning (ML) in chemistry with a focus on neural networks (NNs). When trained on QM/MM reference data, ML models enable dynamical simulations of both ground- and excited-state potential energy surfaces.[220,221] For instance, the photodynamics of the amino acid tyrosine have been investigated using deep neural networks.[222] Similarly, the dynamics of hydroxyl radicals on ice surfaces were explored using neural-network potentials.[223] A comprehensive overview of machine learning approaches for excited-state simulations is provided in Ref. [224]. Molecular dynamics programs, such as SHARC, when coupled with machine learning approaches, enable efficient simulations of photodynamics and surface hopping, respectively, as seen in the SchNarc and S_{PAI}NN frameworks.[225,226] One remaining drawback of these methods is their high computational cost, which remains comparable to that of semi-empirical approaches such as DFTB. Ref. [227] demonstrates that a trajectory surface hopping approach combining semi-empirical density functional tight binding (DFTB)

with neural networks, used to predict Coulomb coupling and excitation energies, can successfully model exciton transfer in light-harvesting systems such as the Fenna–Matthews-Olson complex in green sulfur bacteria.

In Ref. [228], ML models trained on carotenoid chromophores in methanol, incorporating artificially added charges, were shown to accurately predict excitation energies ω in protein environments, successfully capturing solvatochromic effects.

Our study builds on a similar concept, aiming to predict excitation energies (ω) and oscillator strengths (f), and thus absorption spectra, of retinylidene proteins using NNs. Photoreceptor systems pose particular challenges due to their large system sizes, the complexity of protonation states, and the intricate electronic structure of their chromophores.[23] To circumvent costly excited-state calculations on full retinal proteins, we instead simulate two isomeric forms of the chromophore (11-cis and all-trans retinal) in various solvents (water, methanol, acetone, etc.), mimicking the protein environment. To account for different counterion configurations, ions were added systematically to the systems.

The geometries are generated from ground-state QM/MM simulations, with the chromophore described at the DFTB3level.[58,59] Snapshots from these trajectories provide input for calculations of spectroscopic properties carried out with two semi-empirical approaches: (i) orthogonalization model 2/multi reference configuration interaction (OM2/MRCI),[54,229,230] known to reliably capture qualitative features of retinal systems,[231] and (ii) long-range corrected (LC)-time-dependent (TD)-DFTB,[61,63,129] which is computationally cheaper but less accurate.

Retinal chromophore geometries, electrostatic potentials (ESPs), and atom types serve as input features for the NNs, while the spectroscopic properties (ω and f) provide the targets. Two NN models are trained: one predicts excitation energies ω and oscillator strengths f directly at the OM2/MRCI-level, while the other is a Δ-machine that learns the difference between OM2/MRCI and LC-TD-DFTB ($\Delta\omega$ and Δf). The predictive performance of these networks is then evaluated using input features from ground-state QM/MM simulations of different retinal proteins. In this way, we assess whether the models can reproduce qualitative trends in absorption spectra without the need for costly excited-state calculations, while also providing insight into the electrostatic interplay between chromophore and protein environment.

6.2 Computational Details

6.2.1 System Preparation

6.2.1.1 Training Data

The retinal structures used in this study are based on coordinates provided in Ref. [232] for the all-trans retinal structure and Ref. [233] for 11-cis retinal. Atomic charges for both isomers were derived using the restrained electrostatic potential (RESP) fitting method, as implemented in *Antechamber*,[121,122] part of the AmberTools software suite.[119] The electrostatic potentials used for the RESP fitting were obtained from electron densities calculated at the HF/6-31G* level of theory using Gaussian09.[123]

Two capping strategies for the retinal molecules were considered, reflecting the fact that in proteins, retinal is covalently bound to a lysine side chain. Since our simulations are intended to generate training data for neural networks that predict excitation energies ω in retinal proteins, retinal-lysine conjugates were constructed. In one approach, only the lysine side chain was included; in the other, additional capping groups (N-methylamide (NHE) and acetyl (ACE) groups) were added to mimic the protein backbone.

The retinal structures were simulated either in vacuum or solvated in various solvents, such as water, methanol, and acetone. For aqueous systems, the TIP3P water model was employed.[34] The topology files for the additional solvents were obtained from the virtualchemistry website,[234,235] which provides directly force field parameters compatible with the GROMACS program package.[124,125]

The simulations were carried out in a cubic box with a minimum solute-to-box distance of 15 Å. Depending on the system, either one or two Cl^- ions were introduced by replacing solvent molecules near the protonated retinal Schiff base ($RSBH^+$), or one Na^+ ion was added either near the $RSBH^+$ or in the vicinity of the β-ionone ring. In addition, four Cl^- and three Na^+ ions were added to each solvent system to modulate the electrostatic environment, mimicking charged amino acids possibly found in a protein bonding pocket (see Table 6.1 for details). To ensure consistent positioning, the distances between the ions and nearby atoms in the retinal structure were restrained to 3 Å using distance restraints applied via the PLUMED plugin (version 2.5.1).[47,117] Periodic boundary conditions were applied throughout the simulations. The Verlet algorithm was used for neighbor searching, with a cutoff distance of 12 Å applied to both neighbor list generation and Lennard-Jones interactions. Long-range electrostatic interactions were treated using the particle mesh Ewald (PME) method.

Table 6.1: Overview of the systems set up for the two retinal structures depending on the solvents (from polar to unpolar) and the combination of ions (total number of different systems in the training dataset: 67). The different ion combinations are pure (no ions added); with one or two Cl^- or one Na^+ near the $RSBH^+$; one Na^+ near the β-ionone ring; or with several Na^+ and Cl^- atoms. X in brackets denotes systems that were later excluded from the training dataset.

	11-cis retinal						all-trans retinal					
	pure	1 Cl^-	2 Cl^-	1 Na^+	Na^+ ring	NaCl	pure	1 Cl^-	2 Cl^-	1 Na^+	Na^+ ring	NaCl
Water	X	X	(X)	X	(X)	(X)	X	X		(X)	(X)	
Methanol	(X)	X	X	X	X	(X)	X	(X)	X			
Ethanol							X	X	(X)	X	X	X
DMSO							X	(X)	X	X	X	(X)
Anisole							X	(X)	(X)		(X)	(X)
Pyridine							X	X	(X)	(X)	(X)	X
THF	X	X	X	(X)	X							
Chloroform	X	(X)	X		(X)							
Isobutan	X	(X)	(X)		(X)	(X)						
Toluene	X	X	(X)		(X)	X	X	(X)	(X)		X	
Vacuum	X						X					

6.2.1.2 Proteins

Our previous studies provide the initial structures and force-field parameters for the proteins: channelrhodopsin 2 (ChR2),[216] bacteriorhodopsin (bR),[231] Chrimson,[236] and animal rhodopsin.[214] Protonation states of titratable amino acids were determined using the H++ web server.[191–194] Crystal water molecules were retained, and the entire systems were solvated with TIP3P water.[34] Mutations were introduced into the Chrimson and bovine rhodopsin structures using PYMOL 2.5,[237] replacing S169A (Serine on position 169 was replaced with alanine) and E181Q, respectively, in both monomers. The CHARMM36 force field[238] is used, which includes parameters for an accurate description of the lipid bilayer.[239] The homodimeric protein is embedded in a 1-palmitoyl-2-oleoylphosphatidylcholine (POPC)-bilayer.

6.2.2 Classical Equilibration

Classical equilibration was performed using the GROMACS program package (version 2020).[124,125]

6.2.2.1 Training Data

Energy minimization was carried out using the steepest descent integrator until the maximum force fell below $1000\,\mathrm{kJ\,mol^{-1}\,nm^{-1}}$, with a step size of 1 fs. Subsequently, NVT equilibration was carried out at 500 K for 10 ns with a time step of 2 fs, using velocity rescaling (v-rescale) for temperature coupling. The elevated temperature was chosen to enhance sampling and promote increased molecular motion. All bonds involving hydrogen atoms were constrained using the LINCS algorithm. Throughout the classical simulations, the retinal geometry was fixed using position restraints, due to known limitations of classical force fields in accurately modeling the hydrogen bonding networks in retinal protein active sites.[216,240–242] For comparison, additional tests were conducted without position restraints on the retinal atoms. Only the NVT ensemble was used, as prior NPT simulations at this temperature led to instability and box expansion during early testing.

6.2.2.2 Proteins

During classical equilibration of the rhodopsin and Chrimson proteins, the atoms of the retinal chromophore and the covalently bound lysine side chain were position-restrained using a force constant of $k = 1000\,\mathrm{kJ\,mol^{-1}\,nm^{2}}$. For all proteins, energy minimization was performed using the steepest descent minimum algorithm until the maximum force was below $1000\,\mathrm{kJ\,mol^{-1}\,nm}$. NVT equilibration was carried out at a reference temperature of 300 K using velocity-rescale temperature coupling and a time step of 2 fs. This was followed by NPT equilibration for 10 ns using the Parrinello-Rahman barostat at a reference pressure of 1.013 bar. The Verlet integrator and periodic boundary conditions were applied throughout.

6.2.3 QM/MM Setup

All QM/MM simulations were performed using semi-empirical method DFTB3,[58] as implemented in DFTB+,[65] with the 3OB-f parameter set, which is an extension to the 3OB parameters yielding better vibration frequencies.[59]

A time step of 1 fs was used for both the training data systems and the protein simulations. Each system of the training set was simulated for 2 ns, with geometries and the ESP saved every 1 ps.

The protein systems were simulated as follows: bovine rhodopsin and its mutant for 1 ns, ChR2 for 6 ns, and bR for 1 ns. Chrimson and its mutant were simulated for 9 ns, but only the final 6 ns were used for analysis because position restraints were gradually removed during the inital phase. Simulations of ChR2, Chrimson, and bR were carried out using the 3OB parameter set.

6.2.4 Calculation of the Excitation Energies

Geometry snapshots from the QM/MM simulations were used to calculate excitation energies ω using two semi-empirical methods: (i) OM2/MRCI,[54,229,230] and (ii) time-dependent DFTB with a long-range corrected functional[61–63,243] (LC-TD-DFTB) using the ob2 parameter set.[129] In both cases, only the retinal chromophore was included in the QM region, for the following reasons: (i) Neural network inputs must be fixed sizes. To ensure consistency between retinal in different environments (solvents vs. proteins), we used the retinal structure covalently bound via a protonated Schiff base to the lysine side chain. For structures where the protein backbone was mimicked with capping groups, a link atom was introduced between the C_α and C_β atoms of the lysine residue. (ii) Although restricting the QM region to the chromophore may reduce the accuracy of absolute excitation energies, due to limited treatment of dispersion and polarization effects,[218,219,244] previous studies have shown that relative excitation energies ω are still described reliably using this approach.[216] The surrounding environment was included as point charges. However, we observed discrepancies arising from how point charges were handled in the calculations, particularly due to erroneous treatment of periodic boundary conditions. The resulting impact on the excitation energies was systematically investigated and is discussed in detail in section 6.3.1.1. After computing spectroscopic properties, we applied a threshold on the oscillator strength f and included only those data points in the training set for which the first excited-state transition was from S_0 into S_1. For the protein systems, only method (ii), LC-TD-DFTB, was applied due to the atom number limitation of molecular mechanics (MM) atoms in the OM2/MRCI methodology.

The BLA is defined as the average difference between the lengths of single and double bonds in the conjugated polyene chain:

$$\mathrm{BLA} = \left| \frac{1}{n_s} \sum_{i=1}^{n_s} d_{s_i} - \frac{1}{n_d} \sum_{j=1}^{n_d} d_{d_j} \right|, \tag{6.1}$$

where n_s and n_d denote the number of single and double bonds, and d_{s_i} and d_{d_i} are the lengths of the ith single bond and jth double bond. The BLA was evaluated during the analysis of the simulations and subsequently correlated with the excitation energies.

Additionally, we verified that the torsional angle deformations and the electrostatic potentials in the training set span at least the range observed in the proteins. To gain insight into the conformational space, we performed principal component analysis (PCA),[245] and uniform manifold approximation and projection (UMAP),[246] based on the neural network inputs: geometries, ESPs and atom types.

All data was processed and visualized with python[134] using the matplotlib,[135] NumPy,[136] pandas,[137] seaborn,[138] sklearn,[245] and MDAnalysis[247,248] libraries. The NNs were set up based on Keras,[249,250] and TensorFlow.[251]

6.2.5 Neural Network Training and Spectral Prediction in Retinal Proteins

To predict excitation energies ω and oscillator strengths f from retinal chromophore configurations, a neural network was trained using structural and electronic input features, consisting of atomic geometries, the atom types, and the ESPs obtained from QM/MM simulations. The targets for training are either (i) excitation energies ω and oscillator strengths f computed at the OM2/MRCI level, or (ii) the difference between OM2/MRCI and LC-TD-DFTB excitation energies $\Delta\omega = \omega_{\mathrm{OM2}} - \omega_{\mathrm{DFTB}}$ and oscillator strengths $\Delta f = f_{\mathrm{OM2}} - f_{\mathrm{DFTB}}$ for the Δ-machine learning approach.

Ten percent of the training data was set aside for testing. Of the remaining 90%, 10% were used for validation during the hyperparameter search, and the network was trained on the remaining data using the mean squared error as the loss function. A leaky softplus activation function was used in the hidden layers, with an alpha parameter of 0.03, while the output layer used a linear activation function. The best model was selected after training for up to 2000 epochs, with early stopping based on a validation loss plateau of 250 epochs (see Table 6.2 for details).

Table 6.2: Neural network configuration with hyperparameter ranges for optimization.

Parameter	Values
Training data percentage	90%
Loss function	Mean squared error (MSE)
Activation function (hidden)	Leaky Softplus
Alpha (Leaky Softplus)	0.03
Activation function (output)	Linear
Maximum epochs	2000
Early stopping patience	250
Neurons per layer	20 to 100 (step 5)
Number of layers	2 to 8 (step 1)
Regularization	L2
Learning rate	1×10^{-3}, 5×10^{-4}, 1×10^{-4}
Epochs per trial	20
Reduction factor	2

The initial neural network trained both excitation energies ω and oscillator strengths f simultaneously. However, this approach resulted in poor performance. Consequently, the model architecture was adapted to predict excitation energies ω and oscillator strengths f independently, which led to a significant improvement in prediction accuracy.

Given the heterogeneity of the training dataset, ranging from variation in solvent environments and ion configurations; the predictive performance was evaluated separately for each system. Training data systems that consistently exhibited poor performance were excluded from the final training set to enhance overall model reliability and generalizability.

The trained neural networks were subsequently applied to the protein systems. Specifically, inverse interatomic distances and ESPs obtained from the QM/MM simulations of the proteins served as input features. Using these inputs, the neural networks were rerun to predict the corresponding excitation energies ω and oscillator strengths f. These predictions were then used to construct absorption spectra for each protein system.

6.3 Results and Discussion

6.3.1 Training Data Generation

The two retinal isomers were simulated in various solvents, each with different nearby ion configurations. An overview of all systems included in the dataset is provided in Table 6.1. Two capping strategies were tested: one including only the lysine side chain, and another incorporating NHE and ACE groups to mimic the protein backbone. Both approaches resulted in comparable system stability and similar trends in excitation energies ω.[205,206] For simplicity, and to avoid the need for linker atoms in the training data, only the lysine side chain-capped systems were retained for further analysis. In the protein QM/MM setup, linker atoms are still used as required.

As described in section 6.2.2.1, classical MD simulations without position restraints on the retinal atoms were performed for comparison. In these unrestrained simulations, the retinal chromophore exhibited structural deformation. Subsequent excitation energy analysis revealed an overestimation of the BLA, which consistently yielded a value of 0.11 Å across all systems in the dataset.[205,206] This result highlights the limitation of classical force fields in accurately modeling the retinal chromophore.[216,236,240–242] To preserve the structural integrity of retinal and allow the surrounding environment to equilibrate realistically, position restraints on the retinal atoms during classical equilibration are therefore essential.

Following the QM/MM simulations, some Na^+-containing systems with the ion located near $RSBH^+$ were excluded from the training set. Visual inspection revealed structural distortions, particularly in nonpolar solvents, caused by the strong, unshielded electrostatic influence of the nearby ion on the retinal. Such distortion is unexpected during ground-state simulations and, in some cases, even resulted in negative BLA values, indicating an inversion of the single and double bond character along the polyene chain. Similar distortions were observed in several Na^+Cl^--containing systems, which were also excluded from the training dataset. Again, that was mostly the case in nonpolar solvents. One possible explanation is the extensive use of position restraints to fix the ion configuration, which may interfere with the natural flexibility of the chromophore and contribute to these deformations. As these distorted structures were not included in the training of the NNs from the beginning, they are also omitted from Table 6.1.

6.3.1.1 Incorrect Treatment of PBC Conditions in Excited-state Calculations

Snapshots from the QM/MM MD simulations were used as input geometries for subsequent excited-state calculations using two semi-empirical methods: (i) OM2/MRCI and (ii) LC-TD-DFTB. In both cases, the QM region included only the retinal chromophore, while the surrounding MM atoms were either treated as point charges or used to compute the electrostatic potential (ESP) at each QM atom. Both approaches are expected to yield equivalent results.

During the excitation energy calculations using method (ii), LC-TD-DFTB, discrepancies were observed due to incorrect treatment of the periodic boundary conditions (PBC). Fig. 6.3(a) illustrates this issue by comparing 20 representative data points from the dataset simulated in pure toluene. Five different approaches were used to calculate the excitation energies:

(a) MM atoms coordinates extracted using MDAnalysis with periodic boundary conditions (PBC) corrections applied

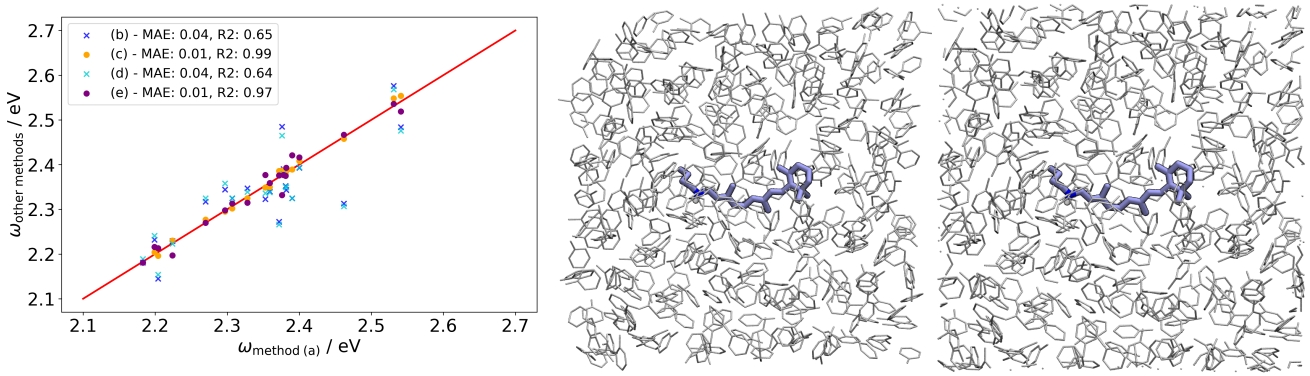
(b) MM atoms coordinates extracted using MDAnalysis without applying PBC corrections

(c) ESP calculated with MDAnalysis considering PBC

(d) MM atoms coordinates obtained on-the-fly during the QM/MM simulation

(e) ESP obtained on-the-fly during the QM/MM simulation

If the excitation energies ω are computed with proper consideration of PBC (method (a), x-axis and (c), orange dots), or the ESP extracted on-the-fly during the QM/MM MD simulation (method (e), purple dots), the results are consistent, with a mean



(a) Comparison ω **(b)** PBC Treatment **(c)** No PBC Treatment

Figure 6.3: Importance of PBC treatment of the MM atoms. (a) Influence on calculated excitation energies ω depending on PBC treatment in various methods compared to method (a), in which PBC corrections are applied using MDAnalysis. 11-cis retinal in toluene (b) with and (c) without treatment of the PBC. The latter reveals broken molecules at the border of the simulation box.

average error (MAE) of less than 0.01 eV. In contrast, excitation energies ω obtained from the point charge file generated on-the-fly (method (iv), cyan crosses) show significant deviation and align closely with those computed without PBC corrections (method (b), blue crosses). This confirms that methods (b) and (d) introduce artifacts due to improper treatment of PBC and should not be used for excitation energies ω calculations.

Figs. 6.3(b) and 6.3(c) illustrate the structural differences arising from neglecting PBC: without correction, the molecules are broken at the box boundaries, leading to inaccurate electrostatics.

Consequently, all subsequent calculations used PBC-consistent approaches: for (i) OM2/MRCI, point charges were extracted using MDAnalysis with PBC (method (a)); for (ii) LC-TD-DFTB, the ESP was calculated with MDAnalysis considering PBC (method (c)).

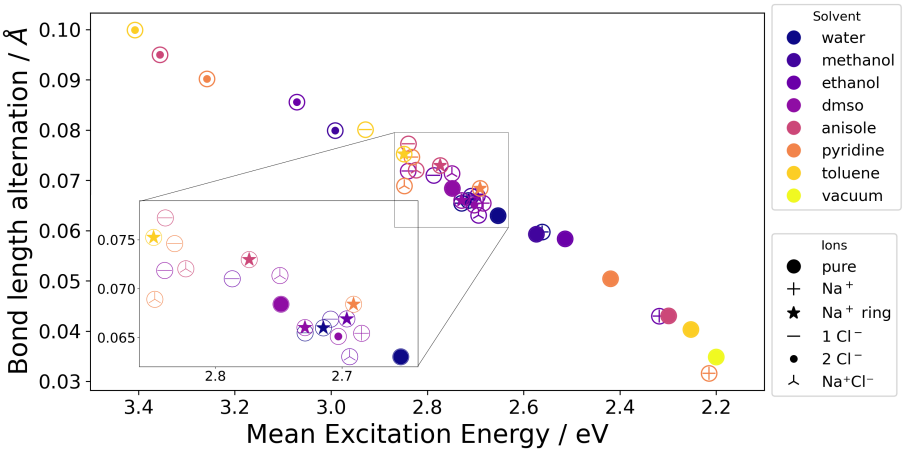
6.3.1.2 Training Data Analysis

Trends of the excitation energies ω for the training datasets

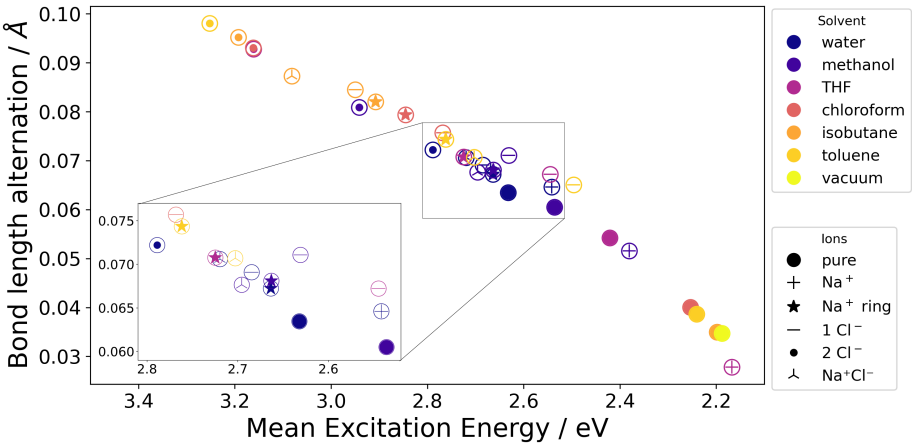
Figs. 6.4(a) and 6.4(b) illustrate the relationship between the mean excitation energies ω_{mean} and mean BLA for all-trans and 11-cis retinal, respectively. In both cases, a clear linear correlation is observed, consistent with earlier findings in Ref.[21], and experimental results linking C=C stretching frequencies to absorption maxima.[252] Among the pure solvents (●), lower polarity environments such as toluene result in lower ω and reduce BLA, whereas more polar solvents like water and alcohols lead to higher excitation energies ω and increased BLA.

The ions near the $RSBH^+$ strongly influence the spectral shift, as expected due to the $RSBH^+$'s high electrostatic sensitivity to its environment.[19,20] Negative Cl^- ions (⊖) cause a blue shift (higher excitation energies), especially when two negative counterions are nearby (⦿), resembling the situation in ChR2, which absorbs high-energy blue light.[216] Conversely, a positive Na^+ ion near the $RSBH^+$ (⊕) induces a red shift, even surpassing that of 11-cis retinal in vacuum, as observed in chloroform. This ionic effect is notably stronger in nonpolar solvents due to the lack of shielding by the solvent. When Na^+ is positioned near the β-ionone ring (⊛), excitation energies ω increase, consistent with the charged and polar residues surrounding the polyene chain and this region in bacteriorhodopsin (bR) and Chrimson that collectively contribute to red-shifted absorption.[209] Systems containing both Na^+ and Cl^- ions show a slight increase in excitation energies, but the effect is minor, likely because ion pairing leads to mutual shielding, reducing their individual influence on the $RSBH^+$.

All observed trends are consistent between the two isomers. Interestingly, the all-trans isomer tends to exhibit slightly higher excitation energies ω compared to 11-cis retinal.



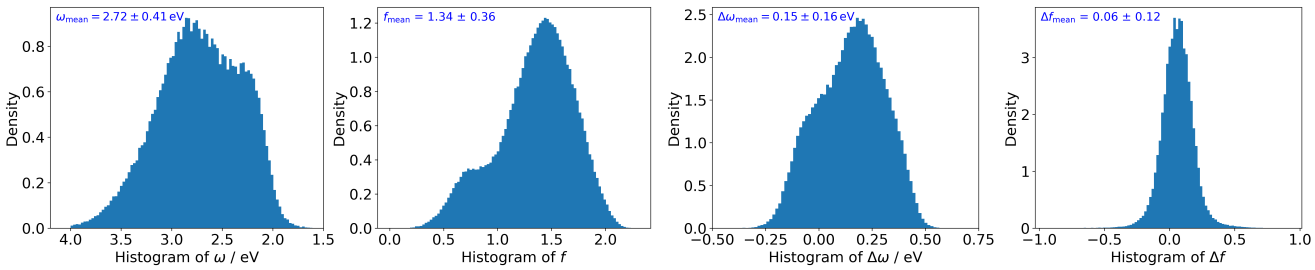
(a) all-trans



(b) 11-cis

Figure 6.4: Linear relation between the calculated excitation energies ω and BLA for training datasets containing (a) all-trans and (b) 11-cis retinal.

Fig. 6.5 shows the distribution of target values used to train the neural networks. For the OM2/MRCI-based model, excitation energies ω span from 1.7 to 4.0 eV, with a standard deviation σ of 0.41 eV. In the case of the Δ-machine model, which predicts the difference between OM2/MRCI and LC-TD-DFTB values, target values range from approximately -0.35 to 0.5 eV. This indicates that LC-TD-DFTB sometimes over- and sometimes underestimates the OM2/MRCI excitation energies. As such, applying a simple constant correction is insufficient, justifying the use of a machine learning-based correction approach as pursued here. Nonetheless, is is noteworthy that in most cases $\Delta_{\mathrm{EE}} \geq 0$, with $\sigma = 0.16\,\mathrm{eV}$, suggesting that LC-TD-DFTB generally underestimates the excitation energy.



(a) Training Data OM2 network

(b) Training Data Δ-machine

Figure 6.5: Histograms of the respective target values, ω and f, for the training data. (a) OM2 values, and (b) differences between methods (i) OM2 and (ii) LC-TD-DFTB. Average and standard deviation depicted in each figure.

Conformational Space of the training dataset and the retinal proteins

The conformational space spanned by the training data was compared to that of the retinal proteins using UMAP and PCA, as shown in Fig. 6.6. The input features are the distances in this case. This analysis included all training datasets and the four proteins, along with their respective mutants.

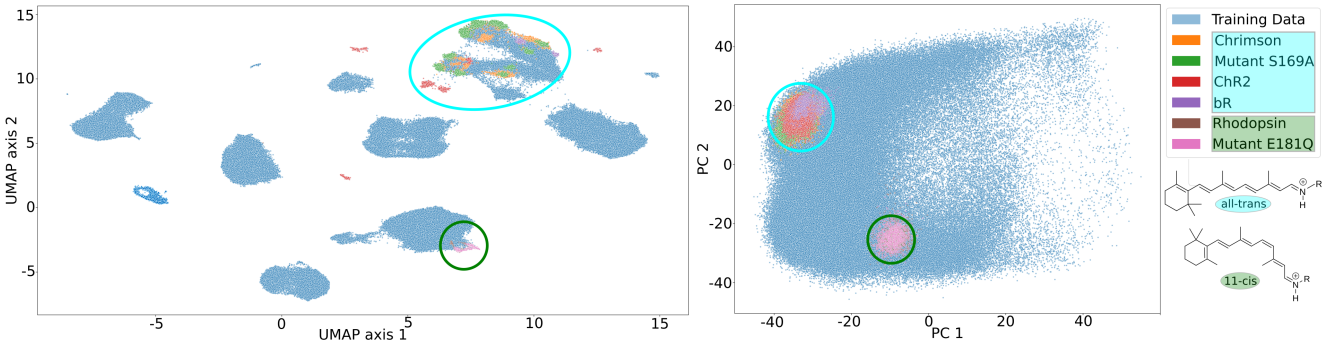
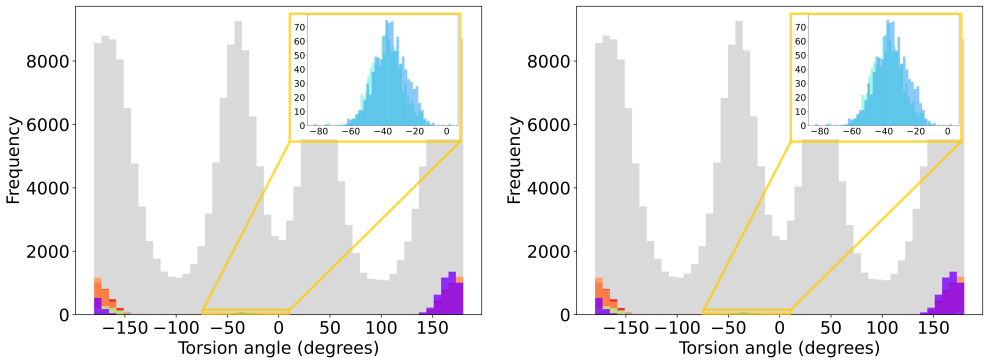


Figure 6.6: Data analysis with UMAP and PCA. Phase space covered by all-trans retinal proteins marked with cyan circles, 11-cis rhodopsin protein, and mutant in green circles. The training data spans over a much larger conformational space than the proteins in both methods.

In both analyses, the training data spans a significantly broader conformational space than the combined protein data. Furthermore, the regions corresponding to proteins with all-trans retinal are generally well separated from those with 11-cis retinal. In most cases, protein data points are embedded within the distribution of the training data, indicating good coverage. An exception is observed for ChR2 in the UMAP representation, where its points appear outside the phase space covered by the training data set. However, this should not be overinterpreted, as the UMAP axes are arbitrarily defined. For PCA, the axes are linear combinations of the original variables, but their physically interpretation remains difficult. Thus, these dimensionality-reduction techniques provide only a qualitative view of the sampled conformational space. A more detailed, feature-level analysis is needed to definitely assess how well the training data represents the proteins.

Torsion angle analysis

Several dihedral angles define the conformation of the retinal chromophore. Fig. 6.7(a) illustrates the distribution of the dihedral angle describing the rotation of the β-ionone ring. The training data (gray histogram) spans the full rotation range, with prominent peaks near ±180° and ±45°, and overlaps more than adequately with the conformational space sampled in the proteins. In the proteins containing all-trans retinal, this torsion angle is found at ±180°, while in the rhodopsins with 11-cis retinal, values cluster around −35°. This reflects known structural differences: in bovine rhodopsin, the polyene chain and β-ionone ring are not planar due to steric hindrance, disrupting conjugation and confining the π-electron system to the



(a) Torsion β-ionone ring

(b) Torsion 11-cis/all-trans isomerization

Figure 6.7: Scan of two torsion angles in the retinal chromophore. (a) Torsion angle at C_6-C_7 bond. (b) Torsion angle at C_{11}=C_{12} double bond. The training data spans the full rotation or a much broader range of the torsion angles, while the proteins are fixed in their respective conformation.

chain. In contrast, microbial rhodopsins accommodate a more planar geometry, allowing the π-system to extend from the polyene chain into the β-ionone ring.[19]

The two retinal isomers (11-cis and all-trans) present in the training dataset can be clearly distinguished based on the torsion angle around the isomerization-relevant double bond, as shown in Fig. 6.7(b). Angles near 0° correspond to the cis-conformation, whereas values around ±180° indicate a trans-conformation. The training data is evenly distributed between both isomers. No full rotation is observed here, which is expected due to the high energetic barrier for isomerization. Upon light absorption, retinal is promoted to the first excited state (S_1), initiating a sequence of events: first, a change in the bond length alternation (BLA) along the polyene chain occurs, followed by isomerization around the C_{11}=C_{12} double bond in animal rhodopsin or the C_{13}=C_{14} double bond in bacterial rhodopsins, and finally relaxation back to the ground state through the conical intersection.[19] Since all training data were generated from ground-state simulations, no isomerization is sampled. In line with expectations, proteins containing all-trans retinal cluster around ±180°, while bovine rhodopsin structures are centered near −20°.

ESP analysis

The electrostatic potential (ESP), calculated on-the-fly during the QM/MM MD simulations, was used as an input feature to the neural network. It implicitly captures the electrostatic influence on each QM atom. Fig. 6.8 shows the ESP values exemplarily for atoms involved in torsional motion, such as C_7/C_8 (rotation around the β-ionone ring), C_{11}/C_{12} and C_{13}/C_{14} (isomerization in animal and microbial rhodopsins, respectively). Additionally, ESP values are shown for the protonated Schiff base nitrogen (N_1) and hydrogen (H_{10}), given their functional importance.

Importantly, the training data spans the full range of ESP values observed in the proteins, indicating good coverage. Carbon atoms tend to exhibit slightly lower ESP values (ranging from approximately -5 to 5 eV), whereas nitrogen and hydrogen atoms predominantly show positive ESP values. Among the proteins, rhodopsins consistently show a shift towards higher ESP values. In contrast, Chrimson, its mutant, and bacteriorhodopsin (bR) cluster closely together. Interestingly, ChR2 shows either intermediate ESP values, falling between the two retinal isomers, or values even lower of those of the other all-trans proteins.

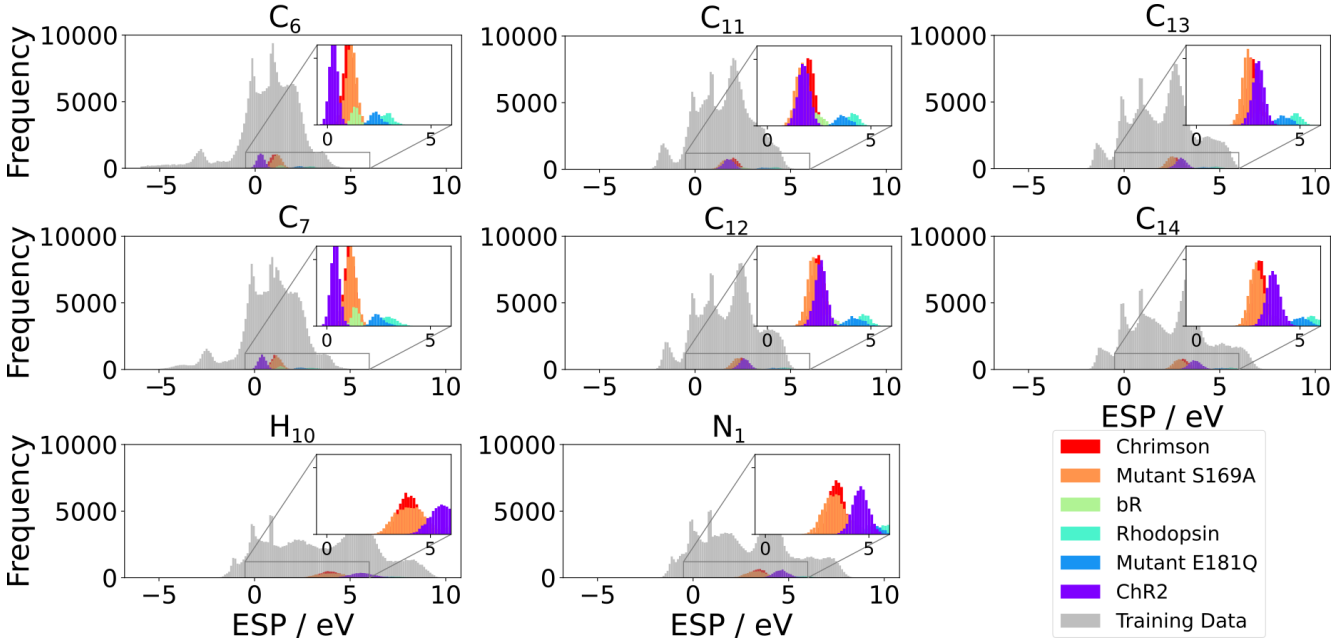


Figure 6.8: Electrostatic potential (ESP) of selected atoms in the retinal. The training data spans a much broader range of ESP values than the proteins.

6.3.2 Neural Networks

The training data described above, comprising quantum mechanical (QM) atom types, geometries, and the ESPs, served as input features for training two types of neural networks. The first model directly predicts excitation energies ω and oscillator strengths f at the OM2/MRCI level of theory. The second model, a Δ-machine learning approach, is trained to predict the difference between the OM2/MRCI and LC-TD-DFTB results. To ensure consistency and reliability of the target values, only transitions corresponding to the first excited state ($S_0 \rightarrow S_1$) were considered, and a maximum threshold ($f \leq 4$) was applied to the oscillator strength to exclude nonphysical transitions.

6.3.2.1 Gradual Improvement of the Networks

Table 6.3 summarizes the performance metrics of the trained neural networks. The initial models (Networks 1) were designed to predict both target properties, ω and f, simultaneously, resulting in a combined Pearson's R^2 score. The OM2 network performed reasonably well, achieving an R^2 of 0.88. In contrast, the Δ-machine learning model showed considerably lower performance, with an R^2 of 0.66. The mean absolute error (MAE) amounted to more than 20 % of the corresponding standard deviation σ for both ω and f. Notably, for f predictions using the Δ-machine, the MAE reached half the value of the σ, indicating limited predictive accuracy.

Table 6.3: Scores of all models, reported as Pearson's R^2, the MAE, and the spectroscopic property relative to the target standard deviation. Network 1 is trained on all data to simultaneously predict both spectroscopic properties. Networks 2 and 3 treat ω and f separately, with Network 3 further excluding the least reliable training sets.

		ω / eV			f		
		R^2	MAE	% of σ	R^2	MAE	% of σ
Networks 1	OM2 network	0.88	0.09	21.2	0.88	0.09	25
	Δ machine	0.66	0.05	29.4	0.66	0.06	50
Networks 2	OM2 network	0.90	0.08	18.8	0.95	0.06	16.7
	Δ machine	0.86	0.04	24.4	0.75	0.04	33.3
Networks 3	OM2 network	0.93	0.07	16.8	0.95	0.06	16.7
	Δ machine	0.90	0.03	20.6	0.78	0.04	33.3

Due to the poor performance of the initial Δ-machine, particularly in predicting oscillator strength, the network architecture was revised to allow separate training of the two target properties (Networks 2). This modification led to a significant improvement in the prediction of ω by the Δ-machine, with the R^2 increasing to 0.86. The performance of the OM2 network also improved slightly, most notably in predicting the oscillator strength f, achieving an R^2 of 0.95.

Due to the heterogeneity of the training dataset, we evaluated the performance of each subset individually by rerunning the OM2 neural network and assessing its prediction accuracy. Fig. 6.9 gives an overview of the R^2 values of the prediction of ω with the OM2 Network 2. Datasets with poor performance ($R^2 < 0.8$, red dotted line) were excluded from further training.

As previously observed during the visual inspection of the training datasets, systems containing Na^+ and Na^+Cl^- ions usually showed the poorest performance. By iteratively retraining and evaluating the performance of the NNs, systems with a performance of $R^2 < 0.85$ were excluded. This was stopped, when excluding additional datasets did not results in improved overall performance. In Network 3, the system performing worst yielded a R^2 value of 0.78. Systems excluded from the training dataset are indicated in brackets in Table. 6.1.

The performance of Networks 3 is presented in Fig. 6.10. Both networks demonstrate further improvement in predicting excitation energies. For the OM2 network, the MAE is reduced to 17% of σ, comparable to values reported in a recent study on carotenoid solvatochromism.[228] While the Δ-machine's prediction of f remains limited (33.3% of the σ), its excitation energy prediction shows significant enhancement.

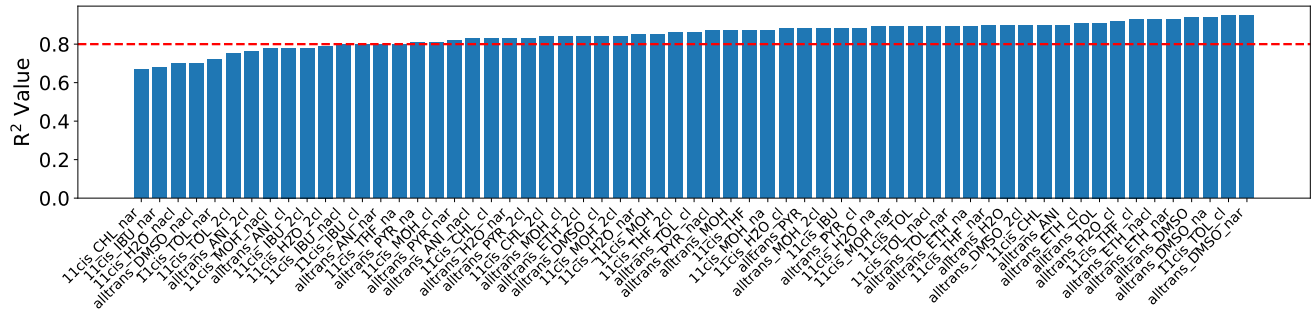
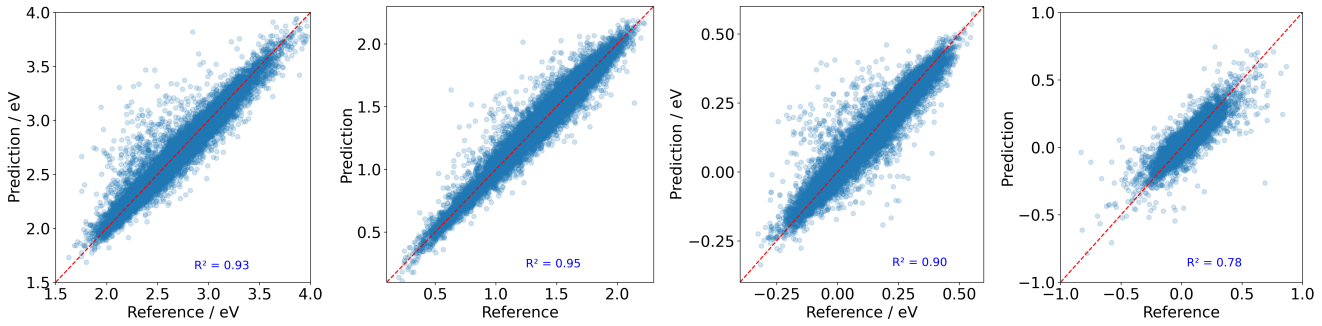


Figure 6.9: R^2 values for OM2 Network 2 predictions of ω. All systems with values $R^2 < 0.85$ (red dotted line) were excluded from further training.



(a) ω_{OM2NN} **(b)** f_{OM2NN} **(c)** $\omega_{\Delta-\mathrm{machine}}$ **(d)** $f_{\Delta-\mathrm{machine}}$

Figure 6.10: Performance of Networks 3, after excluding training data with worst performance. Scatter plots with reference vs. prediction for (a) excitation energies ω of the OM2 network, (b) oscillator strengths f of the OM2 network, (c) ω of the Δ-machine, and (d) f of the Δ-machine. Prediction accuracy is enhanced significantly.

6.3.2.2 Prediction of Spectroscopic Properties of Retinal Proteins

The primary goal of our study was to predict ω and f of retinal proteins to compute absorption spectra at the OM2/MRCI level of theory. While OM2/MRCI is more computationally demanding than LC-TD-DFTB, it provides more reliable results, particularly when comparing to more extensive density functional theory (DFT) methods using the hybrid ωB97XD functional. Although wavefunction-based methods like SORCI[88] outperform these approaches in accuracy, their high computational cost makes them unsuitable for retinal proteins such as ChR2 and Chrimson, which require extensive sampling of their flexible active sites to converge absorption maxima.[216,236]

OM2/MRCI also imposes limits on the total number of MM atoms, preventing full calculations of large proteins such as Chrimson and rhodopsin. For example, tests excluding the POPC membrane in ChR2, bR, and Chrimson resulted in an overestimation of the opsin shift between the retinal in vacuum and in protein.[236]

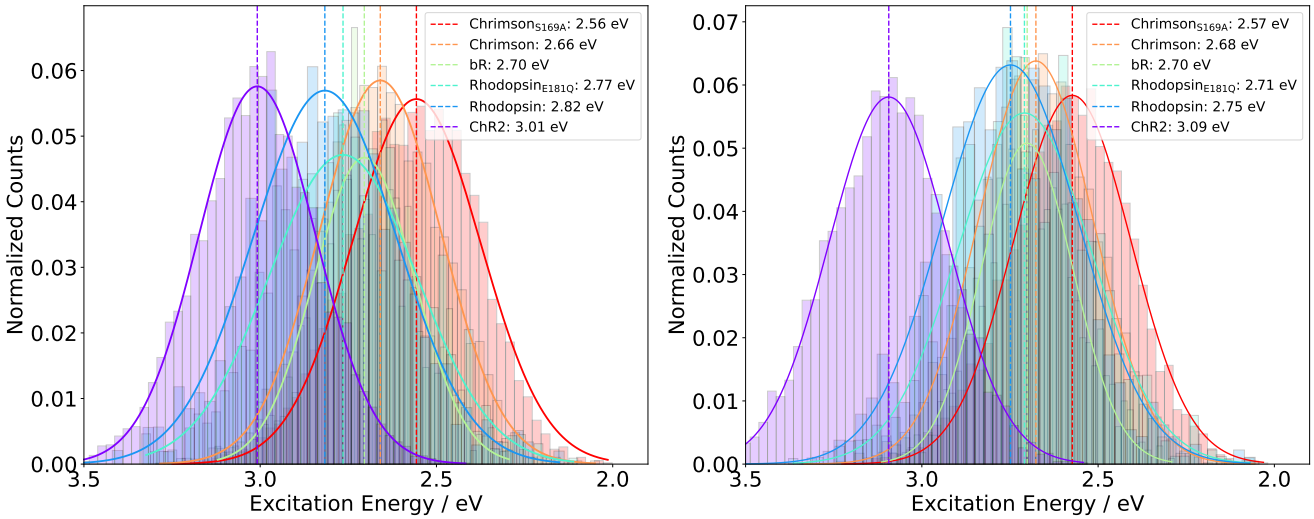
In contrast, neural networks offer a significant advantage in speed and scalability: after the initial preparation of the training data, no further expensive excited-state calculations are needed. Moreover, by using ESPs as input instead of explicit MM point charges, the networks eliminate restrictions on system size, enabling accurate predictions for large protein environments without sacrificing computational efficiency.

Fig. 6.11 shows the absorption spectra computed from the ω and f predicted by the neural networks, while Table 6.4 presents the retinal-protein shifts for context. Both networks qualitatively reproduce the expected ordering of retinal protein absorption: ChR2 exhibits a pronounced blue shift (i.e., higher-energy absorption), whereas bR, Chrimson, and its red-shifted mutant absorb at lower energies. The spectral shifts Δ^* are reported relative to bR, a common practice to mitigate systematic errors arising from the methodology.[231,236] While the Chrimson shifts relative to bR align well with experimental values, the shifts between the rhodopsin proteins and bR are significantly underestimated, particularly by the Δ-machine. The ChR2 shift is also underestimated, though less so than in the rhodopsin cases.

Table 6.4: Predicted values for the retinal proteins. Reruns performed with Networks 3. Relative shift Δ* to bR is given for comparison purposes. Experimental literature values for retinal in gas phase are reported between 610 to 620 nm, approximately 2.00 eV.[253,254]

Protein	Ex. Abs. Max.			Predicted Energies			
				OM2 Network		Δ-machine	
	nm	eV	Δ* in eV	eV	Δ* in eV	eV	Δ* in eV
$\text{Chrimson}_{\text{S169A}}$	608	2.04	-0.14	2.56	-0.14	2.57	-0.13
Chrimson	590	2.10	-0.08	2.66	-0.04	2.68	-0.02
bR	570	2.18	-	2.70	-	2.70	-
$\text{Rhodopsin}_{\text{E181Q}}$	508	2.44	0.26	2.77	0.07	2.71	0.01
Rhodopsin	500	2.48	0.30	2.82	0.12	2.75	0.05
ChR2	470	2.62	0.44	3.01	0.31	3.09	0.39

*Shift between respective protein and bR as reference.



(a) Absorption spectra prediction OM2 NN

(b) Absorption spectra prediction Δ-machine

Figure 6.11: Prediction of spectroscopic properties ω and f yields absorption spectra of several retinal proteins. Rerun of the (a) OM2 network and (b) Δ-machine. Comparison to experimental values shows qualitative reproduction of the color-shifts in both networks.

We conclude that, as proof of principle, our approach performs well and enables the correct ranking of retinal proteins according to their color-shifting abilities. However, the performance of the Δ-machine remains limited, and the overall methodology is constrained by the accuracy of the underlying method used to generate the target values.

6.4 Conclusions

A large and diverse set of training data was generated to provide both input features and target values for neural network training aimed at predicting spectroscopic properties. Evaluation of this data confirmed broad coverage of conformation space, including relevant torsion angles and electrostatic potentials, effectively spanning the phase space occupied by the retinal proteins under investigation. Consequently, the neural networks operate within the bounds of the training data, avoiding extrapolation, a process known to impair predictive reliability.

Assessment of the target values, particularly excitation energies, confirmed consistency with experimental trends, supporting the reliability of the underlying QM methods. Color tuning in retinal proteins, i.e., the broad coverage of absorption across the visible spectrum, arises from electrostatic and dispersion interactions in the protein binding pockets as well as counterion configurations. These effects were successfully mimicked by simulating retinal in different solvents and ion arrangements, yielding the expected behaviors: excitation energies correlate linearly with the bond length alternation; nearby negative charges induce hypsochromic (blue) shifts, especially in nonpolar solvents, while positive charges cause bathochromic (red) shifts, though sometimes accompanied by retinal deformation. We are encouraged to find that the intricate chromophore-protein electrostatic interplay can be effectively reproduced by solvent and ion configurations, achieving broad absorption while greatly simplifying the simulation setup compared to full protein systems.[19–21,216]

The performance of the neural networks was enhanced by modifying the architecture to learn excitation energies ω and oscillator strengths f separately, and by excluding training datasets that exhibited comparatively poor predictive performance. Training ω and f independently enabled the networks to better capture the distinct underlying patterns and complexities associated with each target, leading to improved prediction accuracy.

As outlined above, we successfully demonstrate the proof of concept for our machine learning-based approach, as it qualitatively ranks a diverse set of retinal proteins according to their color-shifting properties. However, limitations remain, particularly in the performance of the Δ-machine networks. Specifically, the opsin shift between retinal in vacuum and within proteins tends to be overestimated. In contrast, the relative shifts with reference to bR are underestimated in the case of bovine rhodopsin and ChR2. Despite these challenges, our approach shows strong potential. The NNs are trained on the smaller training data systems (around 12.000 atoms, roughly one fifth of the size of the retinal proteins tested in this work), where resource-intensive excited-state calculations are feasible. Crucially, we avoid the technical bottleneck of the OM2 method, which imposes an upper limit on the number of MM point charges. By using ESPs as input features instead, this limitation is removed, enabling the trained model to be applied to much larger protein environments.

Further developments could explore whether absorption spectra derived from spectral density analyses provide improved results. Additionally, performing OM2/MRCI calculations on proteins small enough to fit within the method's size restraint, such as ChR2 and bR, would allow for a direct quantitative evaluation of neural network performance. Based on these outcomes, it may be worth considering a switch from OM2/MRCI to more accurate theoretical methods, such as DFT using the ωB97X functional or even wavefunction-based approaches like SORCI. However, this would significantly increase the computational cost required to generate the training data.

Altogether, this framework paves the way for efficient, machine learning-based exploration of spectroscopic properties in retinal-containing photoreactive enzymes.

7 Active Site Structure of Chrimson

Chapter 7 is reproduced from Ref. [236] with permission from the PCCP Owner Societies:

- Spies, Katharina; Bold, Beatrix; and Elstner, Marcus; Complex Active Site Structures influence Absorption Spectrum of Chrimson Wild Type and Mutants., Phys. Chem. Chem. Phys. 27. 2025. pp. 15544–15556.

Author Contributions:
Beatrix Bold set up the Chrimson wild type (wt) structure and models I-III, and performed classical equilibration and simulations, as well as QM/MM optimizations. Katharina Spies prepared the mutant structures and the remaining wt models IV-VIII, conducted the QM/MM simulations, performed excitation energy calculations, and carried out structural analysis during her masters thesis.[255] The preparation of the manuscript for publication was carried out by Katharina Spies and required additional simulations of the counterion mutants, and especially further excited state calculations of all models. Furthermore, analysis that incorporated hydrogen bonding, torsion angles, and the correlation of excitation energies with structural motifs in the active site was conducted. Katharina Spies and Marcus Elstner made equal contributions to the writing of the text.

7.1 Introduction

Rhodopsins are light-sensitive proteins that enable organisms to detect and respond to light. Among them, microbial rhodopsins include channelrhodopsins (ChRs), which are light-gated ion channels that conduct ions across cell membranes upon activation. ChRs are central to optogenetics, where they are used to modulate neuronal activity with high temporal and spatial precision by controlling membrane potential using light.[20,256,257]

Channelrhodopsin 1 and 2 (ChR1, ChR2) from *Chlamydomonas reinhardtii* are activated by blue light.[258,259] For optogenetic applications, however, red-light-activated ChRs are preferred due to deeper tissue penetration.[209] Chrimson, a red-shifted ChR with an absorption maximum at 590 nm, was identified in the algae *Chlamydomonas noctigama*.

Although far red absorbing rhodopsins such as neorhodopsin (NeoR)[211] and bestrhodopsin (BestR)[210] have recently been discovered, Chrimson still remains the most red-shifted cationic ChR in nature. Therapeutic applications of Chrimson include the restoration of sight[7–9] and hearing,[10–13] or the treatment of neural disorders.[14,15] Important features of Chrimson are a high proton selectivity and rapid pH-dependent photocurrent kinetics.[260]

Like other ChRs,[19] Chrimson features seven transmembrane (TM) α-helices with the N-terminus facing outward and the C-terminus facing inward. Its overall structure resembles C1C2, a chimera of ChR1 (helices 1-5) and channelrhodopsin 2 (ChR2) (helices 6-7). While Chrimson shares a similar ion gate architecture with other ChRs, its retinal binding pocket contains amino acids resembling those in bacteriorhodopsin (bR), absent in blue-absorbing ChRs. Substituting these residues with those from C1C2 induces a blue shift in absorption.[209] Similar effects are observed in bR and sensory rhodopsin II (SRII),[21,261] where specific residues modulate the polarity distribution around the β-ionone ring, influencing the color tuning.

In Chrimson, several glutamic acid residues form key components of the ion pore (see Fig. 7.1): E124 and E125 at the inner gate, E132 and E300 at the central gate, and E139 and E143 at the outer gate. The active site is located at the center of each monomer and includes retinal, covalently bound to TM7 via a lysine residue (K299), along with two counterions, E165 and D295. The protonated retinal Schiff base ($RSBH^+$) and its immediate environment, particularly the counterions and nearby water molecules, are critical factors for the opsin shift.[209] This has been well established in various rhodopsins[19,212] and

recently reaffirmed in the near-infrared absorbing rhodopsins NeoR[211,262] and BestR,[210] both of which feature a counterion triad. In NeoR, a single negative charge near the $RSBH^+$ (one counterion is deprotonated) appears to cause the narrow, red-shifted absorption.[211] Similarly, BestR's red-shifted absorption maximum at 661 nm is attributed to two protonated counterions, leaving one effective negative charge near the chromophore.[210] A key contributor to Chrimson's red-shifted absorption is the unique protonation state of the counterions; unlike other ChRs, one remains protonated. Experimental evidence points to E165 as the protonated counterion,[209,260,261] with its close proximity to D295 enabling strong hydrogen bonding. Chrimson's absorption spectrum is pH dependent, as in ChR1;[263] titration experiments show that deprotonation of both counterions induces a hypsochromic (blue) shift.[209]

Combined hybrid quantum mechanics/molecular mechanics (QM/MM) methods have been widely used to gain detailed insights into the structure and function of photoreactive proteins, effectively complementing experimental observations. The status quo of these methods as well as current challenges have recently been reviewed, providing a comprehensive picture of the capabilities of these approaches.[23] In this study, we apply QM/MM methods to investigate key structural features of wt Chrimson, specifically the active site structure and the gates, as well as several mutants in these regions. For a considerable number of photoreactive proteins, the standard procedure involves initial equilibration of the overall system using molecular mechanics (MM) force fields, followed by a QM/MM geometry optimization with focus on the active site structure. This optimized structure can then serve as the basis for computing spectroscopic properties or initiating simulations of chromophore excited-state dynamics. While the active sites of these proteins are not static, they often exhibit stable ground state conformations, where thermal fluctuations average out to structures that closely resemble the QM/MM-optimized geometry. This behavior is observed, for example, in bR and proteorhodopsin (ppR), where the distances between active site residues result from such a unique global energy minimum.

The active site structure of ChRs, as previously discussed for ChR2,[216] appears to challenge such a procedure. In the case of Chrimson, we encounter several modeling challenges that are addressed in this study. The key structural features involve

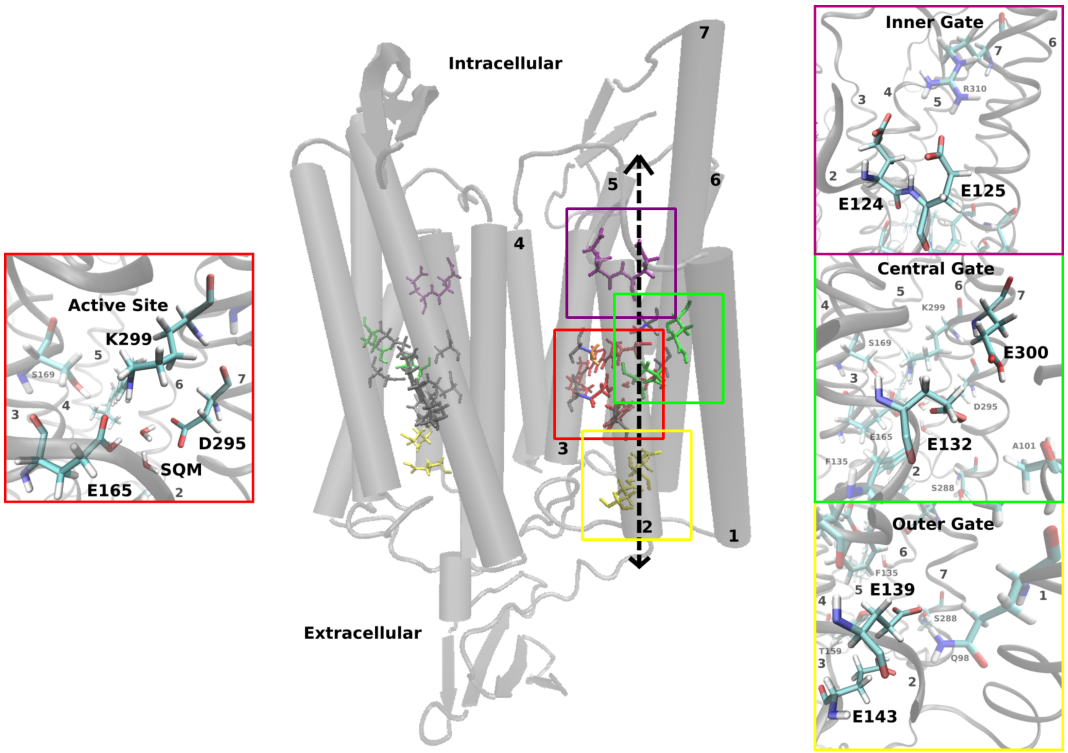


Figure 7.1: Structural model of the Chrimson dimer consisting of two homomonomers with seven transmembrane α-helices each and retinal in the active site; zoom on molecules in the active site (red box); and the inner (purple box), middle (green box), and outer gates (yellow box) with the position of the protonatable glutamates.

extended, hydrogen-bonded networks that exhibit considerable flexibility and support multiple hydrogen bonding patterns. As a result: (i) the protonation states are difficult to determine, (ii) an equilibration of the overall system with MM force fields may lead to unfavorable starting structures of the QM/MM methods,[216] (iii) extensive QM/MM simulations on the ns-timescale are required to capture relevant dynamics and structural heterogeneity. However, such simulations are computationally prohibitive for density functional theory (DFT) or ab initio methods, particularly when very large quantum mechanical (QM) regions are involved. Additionally, the retinal chromophore itself presents a significant challenge for excited states methods.[231,244]

In a first step, to address the uncertainty in protonation states and hydrogen bonding patterns, we constructed several structural models and investigated their stability. These models were further evaluated by computing absorption spectra, also including important mutants. Our results reveal a highly flexible active site that exhibits multiple hydrogen bonding patterns, with their relative occurrence varying depending on the protonation states and specific mutations. This structural flexibility, which is characteristic for ChRs,[21,216,217] is identified as a central factor contributing to the unique red shift absorption in Chrimson.

7.2 Computational Details

7.2.1 Chrimson Models and System Optimization

The Chrimson models used in this study are based on the X-ray structure of wt-Chrimson (PDB ID: 5ZIH). Missing residues not visible in the crystal structure were added using the Swiss model.[190] The resulting monomer was used to construct the initial structure of a Chrimson dimer. The dimer was embedded in a 1-palmitoyl-2-oleoylphosphatidylcholine (POPC) lipid bilayer and solvated by water molecules in addition to those in the crystal structure, as shown in Fig. 7.2.

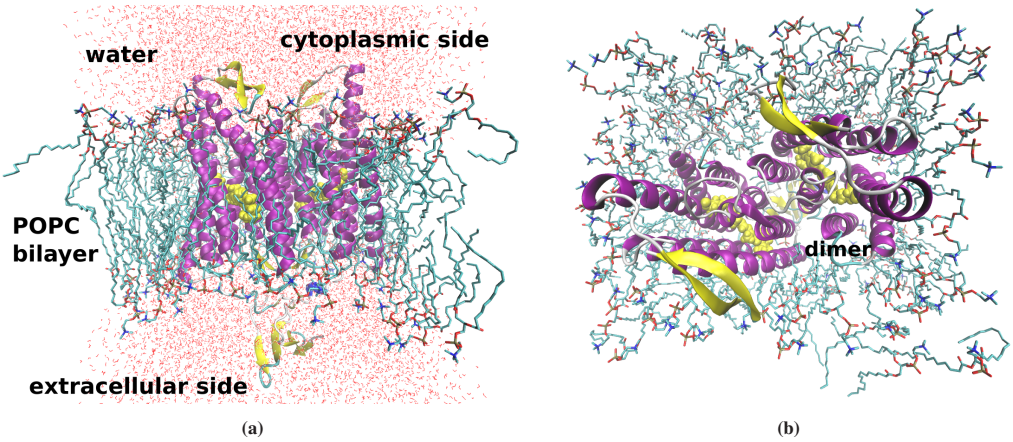


Figure 7.2: Model of Chrimson: A protein dimer (purple) is inserted into a POPC lipid bilayer and surrounded by water molecules on both the cytoplasmic and extracellular sides. The retinal chromophores (yellow) are covalently attached to K299 side chains. (a) Front view; (b) Top view, rotated by 90°.

Experimental studies[209,260,261] suggest that the $RSBH^+$ counterion E165 is protonated, which is one of the main factors responsible for the red shift. However, the protonation state of the $RSBH^+$ counterions (E165 and D295) was modeled as follows: (i) residue E165 protonated and D295 deprotonated and (ii) residue E165 deprotonated and D295 protonated; in order to obtain a comprehensive theoretical study.

Further protonation states of titratable amino acids were not determined by experimental studies, so additional PropKa[264–266] and MCCE[267–273] analyses were performed. Protonation states of histidines were determined by comparison with ChR2. We decided on eight models, see Table 7.1, with different protonation states of the glutamates located in the putative ion transport pathway, based on the calculated pk_a values and experimental suggestions made by comparing the absorption spectra of several Chrimson mutants.[209,261] In model I, all glutamates are deprotonated, creating a highly negatively charged putative ion pathway. Protonation is then gradually increased. Models IV and V are based on protonation states suggested by MCCE and PropKa, respectively. Models VI-VIII build on models III-V by additionally protonating E139, resulting in all glutamates in the central and outer gate being protonated in model VIII.

The mutants were constructed using the same procedure as described for the WT. The respective amino acids were substituted with the program PYMOL 2.5.[237] and the rotamer with the highest probability and least structural disruption was selected.

7.2.2 System Equilibration, Optimization and Production

The simulation protocols are similar to those used in our previous publications on retinal proteins,[215–217,231,241,274,275] using classical molecular dynamics (MD) simulations for the system preparation and for the QM region DFTB3/3ob.[58,59]

Table 7.1: Chrimson models based on gradual protonation of glutamates in the putative ion pathway.

Model	Protonated glutamates			
	E143	E132	E300	E139
I	-	-	-	-
II	x	-	-	-
III	x	x	-	-
IV[a]	-	x	x	-
V[b]	x	x	x	-
VI	-	x	x	x
VII	x	x	-	x
VIII	x	x	x	x

[a]suggested by MCCE; [b]suggested by PropKa

7.2.2.1 Classical Equilibration

Classical MD simulations are unable to describe the complex hydrogen bonding network in the active site of retinal proteins.[216,240–242] In order to generate starting structures for the QM/MM simulations, we performed different classical equilibration protocols with position and/or distance restraints.The equilibration protocols also differ depending on the protonation state of the counterions. All classical MD simulations were performed using the CHARMM36 force field,[238] which includes parameters for an accurate description of the lipid bilayer.[239] The TIP3P water model[34] was used to describe the solvent.

7.2.2.2 QM/MM Simulation

The QM region consists of the retinal (K299), the side chains E165, D295 and F135, and water molecules (hereafter referred to as SQM) in close proximity to the active site. Criteria for the selected molecules in the QM region were derived from former studies on retinal proteins,[216,231] in which the amino acids in the homologous positions were selected. The selection is based on the main residues constituting the active site and this large QM region is necessary to obtain converged structures.[215,216,241] Additionally, their properties played a major role, such as the electrostatic influence of the counterions E165 and D295 or the shielding from bulk water of F135.[209] The oxygen atoms of the QM water molecules were restrained with a force of $500\,\mathrm{kJ\,mol^{-1}\,nm^{-1}}$ to keep them close to the active site and prevent the exchange of QM and MM water molecules. QM/MM structure optimizations, based on the classically equilibrated NPT structures, provided only a partial view of the Chrimson active site, highlighting the need for QM/MM MD simulations to achieve a more comprehensive analysis.

The QM region is the same as that used for the QM/MM optimization. Different simulation protocols were also used depending on the model and distance restraints that were introduced in the classical equilibration were gradually removed to obtain smooth and well equilibrated QM/MM MD production runs. All models were simulated for a minimum of 1 ns. A more detailed analysis was performed for wt-Chrimson models IV and V, for which simulations with E165 protonated in the initial structure were run three times with the same starting structure and different velocities for a total simulation time of 6 ns. Residue S169 was additionally included in the QM region because it strongly influences the polarity and hydrogen bonding patterns in the active site. All Chrimson mutants were simulated with the protonation states of model V, and otherwise the simulations were performed as described above.

For comparison purposes, QM/MM simulations were performed for ChR2[217] and bR[274] with the initial structures and force field parameters taken from our previous studies. Simulation parameters are identical to the Chrimson models. All QM/MM optimizations and simulations were performed using the DFTB3/3ob parameter set[58,59] for the QM region, as implemented in the GROMACS package (version 2020).[124,125,234,276,277] The remainder of the system was modeled using the CHARMM36[238] force field. DFTB3/3ob offers a favorable balance between accuracy and computational efficiency for modeling rhodopsins,

particularly in capturing complex hydrogen-bonding networks.[216,241,242] The accuracy of DFTB/3ob is comparable to that of full DFT calculations with medium-size basis sets.[58] This is supported by: (i) its reliable performance in describing hydrogen-bonding interactions and large water clusters as reviewed in Ref. [278] (Chap. 7.2) (ii) its accurate estimation of proton affinities for a wide range of organic molecules, relevant for the description of relative energies of proton donor-acceptor complexes (iii) and its ability to reproduce proton transfer barriers,[58,59] which are essential for evaluating reaction kinetics. DFTB3/3ob has been shown to achieve accuracy close to that of B3LYP with medium-sized basis sets across these relevant properties, indicating its ability to resolve the relative occurrence of hydrogen bonded motifs as reported in this manuscript, similar to those described in our earlier work.[216,241,242]

Distance restraints on the atoms in the active site were applied with the PLUMED library (version 2.5.1).[47,117]

7.2.2.3 Structural Analysis

The trajectories of the QM/MM MD simulations were used to perform a structural analysis. Emphasis lied on hydrogen bonding patterns and distances in the active site and the gates of the ion conductive pathway. The analysis was performed using the visualization program VMD[279,280] and the open-source, community-developed PLUMED library,[47] version 2.5.1.[117] For the hydrogen bond analysis the following criteria were used: (i) the angle between the H^+ donor N of $RSBH^+$ and the acceptor OE1/OE2 of E165 or OD1/OD2 of D295 ($^+HN\cdots O$) is less than 30°, and (ii) the distance between N and O is less than 3.5 Å.

All data was processed and visualized with python[134] using the matplotlib,[135] NumPy,[136] pandas,[137] and seaborn[138] libraries.

7.2.3 Excitation Energies

The ground state dynamics of the Chrimson active site were described by the DFTB3/3ob method within a QM/MM framework. Excitation energies were calculated in a subsequent step for the respective structures. Only the retinal (residue K299) was part of the QM region, whereas the rest of the protein was modeled by classical MM point charges. This restriction of the QM region to the chromophore and its covalently linked lysine side chain reduces computational cost and facilitates consistent comparison of the relative excitation energies across different proteins, models and mutants. While this simplification may limit the accuracy of absolute excitation energies, due to the exclusion of nearby residues that can participate in charge transfer, leading to blue-shifted energies, these effects are largely systematic within a given protein. Moreover, although a more extensive QM region and explicit treatment of polarization and dispersion would improve the absolute value,[218,219,281] such corrections introduce uniform shifts that cancel out when comparing relative excitation energies. As a result, using only the chromophore in the QM region yields realistic excitation energy trends, as shown in previous studies on rhodopsins.[216]

Excitation energies were calculated with the time-dependent generalization of density functional tight binding (DFTB) using a long-range corrected functional (long-range corrected (LC)-time-dependent (TD)-DFTB[61–63,243] with ob2 parameters[129]) for all QM/MM MD sampled structures (6000 snapshots for each model and mutant, respectively). In the following, LC-TD-DFTB/ob2 will be abbreviated by LC-TD-DFTB. The electronic parameters differ to those used in Ref. [231, 243] and match the ones used in Ref. [282, 283]. The self-consistent charge (SCC) convergence criteria was set to 1×10^{-8} electrons, the range separation parameter to $\frac{0.3}{a_0}$ and the threshold of the screening method is set to 1×10^{-16}. LC-TD-DFTB is computationally very efficient and therefore suitable for the calculation of absorption spectra, where a large number of calculations is required.[231] For comparison purposes excitation energies of two other rhodopsin proteins were computed: bR and ChR2. The structures used for QM/MM simulations were based on previous publications,[217,231] and the snapshots (1000 snapshots in the case of bR and 6000 snapshots for ChR2) used for the excited state calculations. In addition, 250 trajectory snapshots were chosen randomly from the QM/MM simulations of each system to calculate excitation energies on the TD-DFT level of theory using the ωB97X functional[133] and def2-TZVP basis set.[284] These calculations were performed using the ORCA program package version 5.0.3.[285,286]

7.3 Results

7.3.1 Computational Model of Chrimson

This section introduces the computational model of Chrimson used to analyze the active site and investigate the effects of mutations in the subsequent sections. An initial set of proposed models was thoroughly evaluated, and most were discarded due to structural inconsistencies. As detailed in methods section 7.2.2.1, special care was taken during the preparation of the configurations. Classical force-fields are limited in their ability to preserve the complex hydrogen-bonding networks characteristic of retinal proteins,[216,240–242] necessitating an equilibration protocol with gradually released distance restraints. The retinal chromophore, E165, D295, F135, and nearby water molecules, denoted as SQM (see section 7.2.2.2) are included in the QM region.

Protonation states of the counterions Experimental studies at different pH values (6, 7.4, and 9) indicate a pH-dependent protonation for the counterions of the protonated Schiff base ($RSBH^+$) and show a blue shift in the absorption spectra. The pK_a of the counterion E165 is 8, resulting in a protonated residue at physiological pH. This is identified as one of the key factors for the red shift of Chrimson,[260,261] as a reduction of the negative charge near the $RSBH^+$ destabilizes the ground state, leading to absorption at higher wavelengths.[19] While K93 in ChR2 is a conserved residue in channelrhodopsins that stabilizes the deprotonated form of the counterion E123 through a salt bridge,[287,288] it is replaced by F135 in Chrimson. The hydrophobic side chain shields the active site from bulk solvent and increases the pK_a of the glutamate, stabilizing its protonated form.[261,289]

We considered two initial protonation states: i) residue E165 protonated and D295 deprotonated, and ii) residue E165 deprotonated and D295 protonated to investigate, which counterion is protonated. The QM/MM MD simulations resulted in the proton to be found in close proximity to residue E165 regardless of the starting structure as shown in Fig. 7.3. To follow the proton transfer events in both initial conditions i) and ii), Fig. 7.4 shows the distances between the counterions to the respective hydrogen atom exemplarily for Chrimson-wt model IV and V. In model ii), proton transfer occurs immediately at the start of the QM/MM MD simulations, followed by a few back-and-forth events until approximately ≈250 ps (see Fig. 7.4(b) and 7.4(d)). After this point, E165 remains stably protonated with no further transfer. In contrast, when E165 is protonated in the initial structure (model i)), no proton transfer to D295 is observed (see Fig. 7.4(a) and 7.4(c)). These results support experimental findings that E165 is the protonated counterion.[209,260,261]

Protonation states of the glutamates in the putative ion pathway
As stated in methods section 7.2.1, a challenge in simulating rhodopsin proteins is the uncertainty in protonation states of titratable amino acids. The ion-conducting pore in Chrimson is closed in the ground state by three gates, that are lined by glutamates, the inner gate (E124, E125), the central gate (E132, 300) and the outer gate (E139, E143).[209] In order to investigate

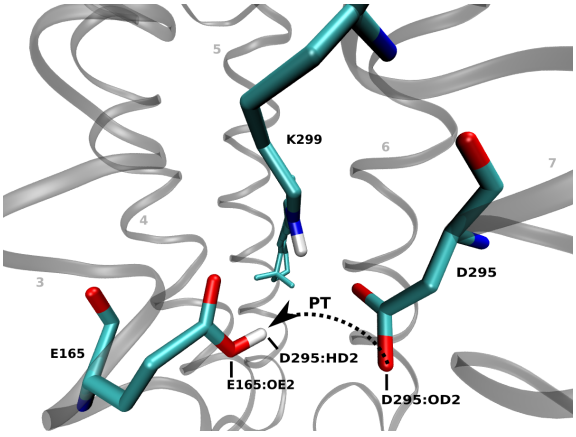
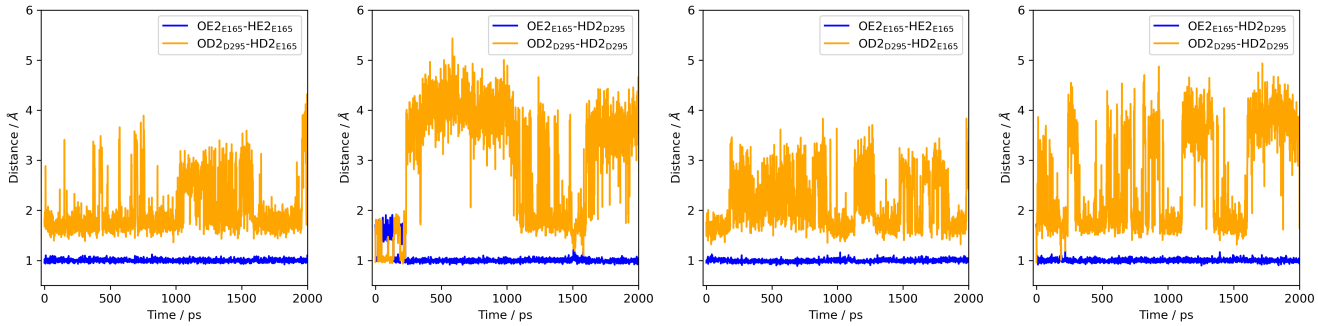


Figure 7.3: The residue D295 is the predominant counterion in Chrimson, while E165 is protonated in the ground state.

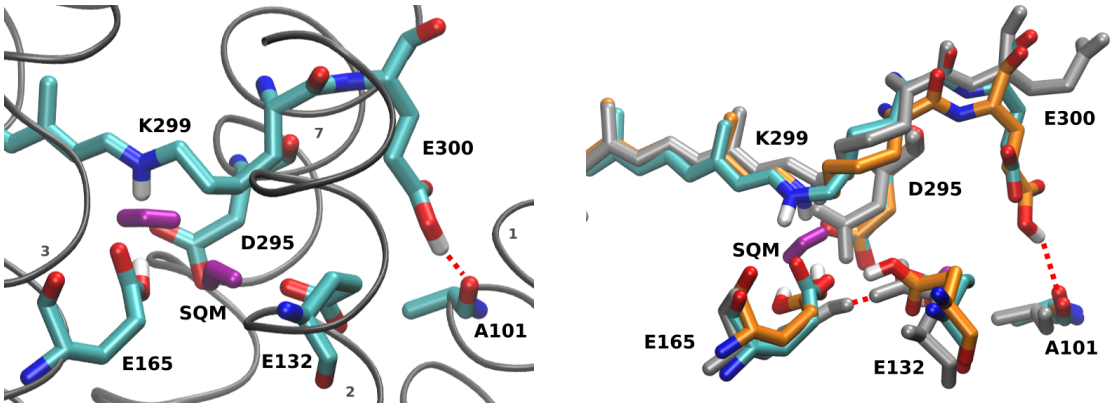


(a) Chrimson-wt_V Model i) **(b)** Chrimson-wt_V Model ii) **(c)** Chrimson-wt_{IV} Model i) **(d)** Chrimson-wt_{IV} Model ii)

Figure 7.4: Distances between the counterions to the respective hydrogen atom ($HE2_{E165}$ in Model i) and $HD2_{D295}$ in Model ii)) to identify proton transfer events.

the protonation states of the amino acids in the putative ion pore, we consider several computational models as summarized in the Table 7.1 in section 7.2.1. Since the glutamates E124 and E125 in the inner gate of Chrimson are homologous to E82 and E83 in ChR2 and both of them are deprotonated in ChR2,[216] we assume these protonation states also for Chrimson.

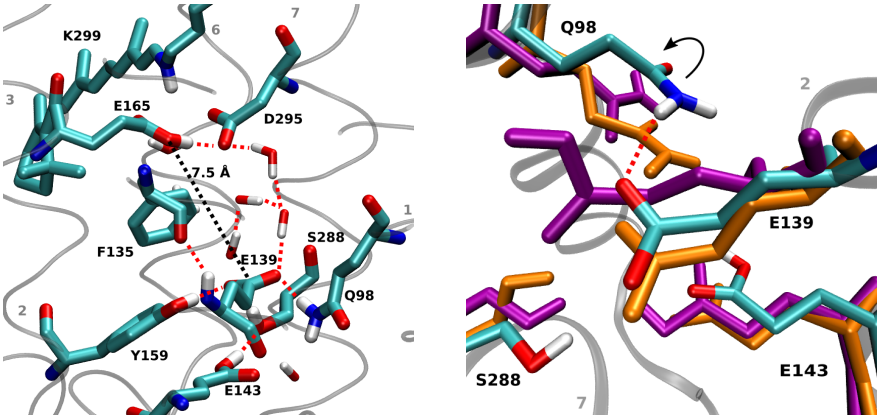
Ref. [209] suggests an ionic interaction between E125 and R310, linking TM2 and TM7. In the central gate of ChR2, there is a neutral N268 and a protonated E90,[216,241,287] while in Chrimson, the central gate contains glutamates E132 and E300, whose protonation turns out to be critical for the stability of the active site (exemplarily shown with wt model V in Fig. 7.5(a)). Our simulations show that deprotonation of E132, as in models I and II, leads to a destabilization of the active site, whereas deprotonation of E300 impedes hydrogen bonding to A101 and therefore a connection of helices 7 and 1 (see in Fig. 7.5(b)). These observations lead to the conclusion that both, E132 and E300, must be protonated, excluding model I to III, to obtain a stable active site and to establish the experimentally observed linkage of helices 1 and 7.[209]



(a) Model V **(b)** Model I (silver), III (cyan) and VIII (orange)

Figure 7.5: Protonation of glutamates E132 and E300 is critical for active site stability. Central gate of QM/MM sampled wt-Chrimson models. Possible hydrogen bonds are indicated by red dotted lines. (a) Stable active site and hydrogen bond between A101 and E300 connecting helices TM1 and 7 due to protonated E132 and E300 (model V). (b) Destabilization of the active site as deprotonated E132 forms a hydrogen bond with E165 (model I, silver structure). No connection of helices TM1 and 7 due to deprotonated E300 (model III, cyan structure). Protonation of both E132 and E300 gives expected results (model VIII, orange structure).

The outer gate in ChR2 is formed by deprotonated E97 and E101,[216] also present in Chrimson: E139 and E143. A complex network of water and hydrogen bonds connects the central gate with these residues (see Fig. 7.6(a)). A stable connection between TM2 and TM3 is observed in all our models as E139 and Y159 are constantly hydrogen bonded during the QM/MM simulations. When E139 is deprotonated, a direct hydrogen bond between E139 and Q98 is formed, as proposed in the crystal structure[209] (see model V (cyan) in Fig. 7.6(b)), thereby connecting helices 1 and 2. In contrast, when E139 is protonated (model VII: purple structure in Fig. 7.6(b)), Q98 turns sideways. This indicates that the protonation of E139 prevents hydrogen bond formation with the amino group of Q98, effectively excluding models VI-VIII from further consideration.



(a) (b)

Figure 7.6: Outer gate structure of Chrimson. (a) No direct interaction of E139 with the active site due to distance (black dotted line), but connection of central and outer gate through water and hydrogen bond network (red dotted lines). (b) Outer gate of wt-Chrimson models. Deprotonated E139 forms a hydrogen bond to Q98, connecting TM 1 and 2 (model IV & V, orange & cyan). Q98 turns sidewards with protonated E139 (model VII, purple).

While the protonation of E139 is structurally critical, the determination of the protonation state of E143 is less straightforward, as models IV and V share similar structural motifs despite the difference in the protonation state of E143. Therefore, in the following sections we present a more detailed analysis of models IV and V.

7.3.2 Active Site of Chrimson

The active site of Chrimson shows a particularly high flexibility, already reported for other channelrhodopsins, such as ChR2,[216] which we will present in detail in section 7.3.2.1. In addition, the unique red light absorption of Chrimson can be related to specific motifs in the active site, as shown in section 7.3.2.2.

7.3.2.1 Structural Motifs

In contrast to bR, where the active site features a stable water-bridged hydrogen bonding network between the $RSBH^+$ and the two counterions D85 and D212,[21,290] Chrimson exhibits a highly flexible active site, fluctuating between distinct structural motives similar to those observed in ChR2.[216] Fig. 7.7 shows the Chrimson active site as determined by X-ray diffraction.[209] The reported interatomic distance between the oxygen atoms of the counterions is 3.3 Å, while the distances from the retinal

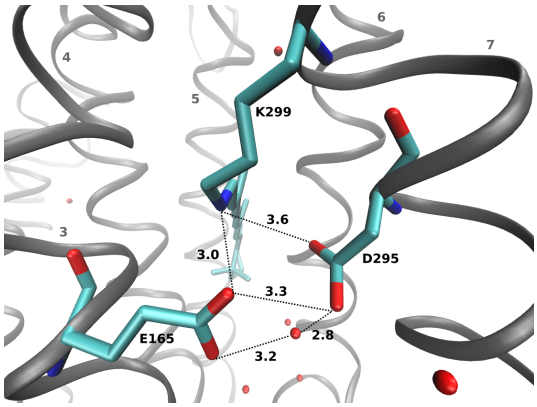


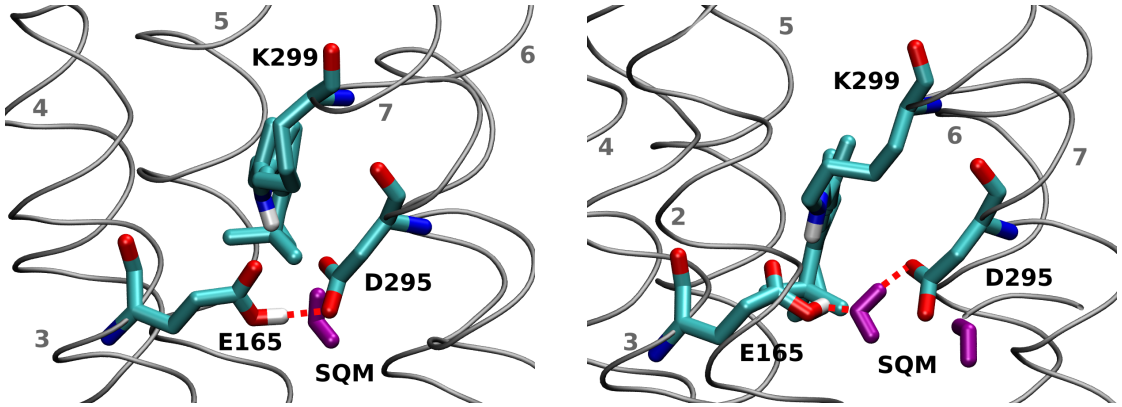
Figure 7.7: The interatomic distance between the oxygen atoms of the counterions in the active site of the Chrimson crystal structure (PDB-ID: 5ZIH)[209] is 3.3 Å, while the distances of the $RSBH^+$ to E165 and D295 are 3.0 Å and 3.6 Å, respectively. A water molecule (red) is in close proximity to the counterions.

nitrogen to the oxygen atoms of E165 and D295 are 3.0 Å and 3.6 Å, respectively. These values appear long, given that E165 interacts with a positively charged retinal and a negatively charged D295, where hydrogen bond lengths are typically shorter than 3 Å.[291] In other retinal proteins, the distance between the $RSBH^+$ nitrogen and the oxygen of a nearby water molecule or counterion commonly falls below <2.9 Å. For example, in bR, the hydrogen bond between retinal and a nearby water molecule is about 2.8 Å; in ppR, this distance is even shorter at 2.6 Å.[21] In ChR2[216] the distance is approximately 2.8 Å, involving either a water molecule or the counterion E123. In the sodium pump *Er*NaR from *Erythrobacter sp. HL-111*, the $RSBH^+$-counterion distance is ≈2.8 Å, while the distance between the two counterions is exceptionally short at just 2.2 Å.[292] We note that the interatomic distances observed in the crystal structure appear to be unusual for retinal proteins. Notably, a water molecule positioned near the counterions in the crystal structure can enter the active site and dynamically rearrange the counterion-retinal complex.

Our simulations reveal two dominant structural motifs in the active site: either the counterion oxygen atoms of E165 and D295 are within a distance of $\leq$ 3.0 Å to each other (Fig. 7.8(a), conformation C1), forming a hydrogen bond; or they are separated by $\geq$ 4.0 Å bridged by a hydrogen-bonded water molecule (Fig. 7.8(b), conformation C2). The average distance between those atoms is 2.8 Å in conformation C1 and 4.2 Å in conformation C2. The 3.3 Å bond length observed in the crystal structure may reflect an average of these two conformations, particularly if they are similar populated. In model IV, C1 is slightly more populated, while in model IV both conformations occur at approximately equal frequency (see Table 7.2). In conformation C1, the proton is not shared between the two counterions: the hydrogen consistently remains closer to E165, and no proton transfer is observed along the trajectory (see Fig. 7.4(a) and 7.4(c)).

Table 7.2: Percentages of the different structural motifs in the active site of Chrimson wt models and mutants

	wt_{IV}	wt_V	S169A	S169D
C1	58.4	49.1	61.4	6.1
C2	41.6	50.9	38.6	93.9



(a) C1 **(b)** C2

Figure 7.8: Two structural motifs in the active site structure are observed along the QM/MM simulations: The counterions are (a) in close proximity and permanently hydrogen-bonded (C1), or (b) a water molecule (SQM) is located between them with hydrogen bonds to both (C2).

E165 side-chain conformation

The orientation of the E165 side chain varies along the trajectory, depending on the torsion angle (C_β-C_γ) (Fig. 7.9), and corresponds to the structural motifs observed in the QM/MM MD simulations. It alternates between a lateral conformation (C1), with a dihedral angle of -120° to -180°, forming a strong hydrogen bond with D295 and a vertical conformation (C2), with a dihedral angle of -80° to -110°, where a hydrogen-bonded water molecule (SQM) bridges the counterions. In both cases, the backbones of E165 and S169 remain hydrogen-bonded. However, an additional hydrogen bond between their side chains stabilizes the vertical orientation of E165, favoring conformation C2.

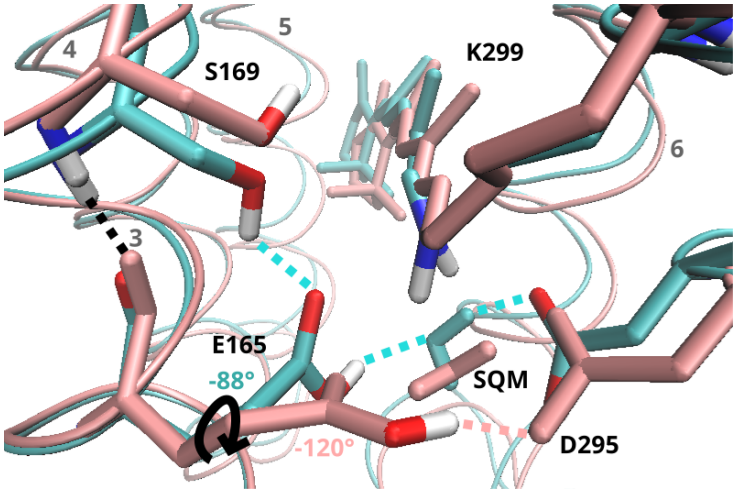


Figure 7.9: The orientation of the E165 side chain changes according to the structural motif (pink: conformation C1; cyan: conformation C2). The black arrow indicates the torsional motion of the E165 side chain from one conformation to the other. Hydrogen bonds are indicated by dotted lines.

RSBH$^+$ hydrogen-bonding patterns

Extensive MD simulations of the ChR2 ground state reveal a highly dynamic active site that samples three main hydrogen bonding patterns: the RSBH$^+$ forms hydrogen bonds with either counterion (E123 and D253) or with a nearby water molecule.[216,217,288] This increased flexibility, compared to bR, is partly due to the substitution of aspartate (D85) with glutamate, whose longer side chain allows greater mobility. A similar dynamic is observed in Chrimson, characterized by water influx and efflux and multiple conformations as described above. Our extended analysis of hydrogen bonding patterns (see Table 7.3) confirms this similarity: hydrogen bonding between the RSBH$^+$ occurs to both counterions and in Chrimson-wt$_V$ also to the nearby water molecule. Interestingly, D295, reported to lie at a distance of 3.6 Å from the RSBH$^+$ in the crystal structure,[209] shows a slightly higher frequency of hydrogen bonding to the RSBH$^+$ than E165. In both models, we observe that not only E165 but also D295 occasionally approaches the RSBH$^+$ within 3 Å, contradicting the crystal structure data.[209]

Table 7.3: Hydrogen bonds between the RSBH$^+$ as acceptor and the counterions (E165 and D295) or nearby water as donor. Criteria to be met for hydrogen bonding: (i) angle$_{^+HN\cdots O} \leq 30°$ and (ii) distance$_{N\cdots O} \leq 3.5$ Å. All results in % of the occurrences along the trajectory

	wt$_{IV}$	**wt$_V$**
RSBH$^+\cdots$E165	23.6	23.0
RSBH$^+\cdots$D295	39.1	30.0
RSBH$^+\cdots$SQM	0.4	14.8
total	63.0	66.3

7.3.2.2 Calculated Absorption Spectra

The excited-state spectra of rhodopsins are highly sensitive to the chromophore's environment: (i) steric interactions can induce a twist in the retinal structure, altering excitation energies, and (ii) hydrogen bonding and Coulomb interactions with charged and polar groups can affect the absorption spectrum drastically. Consequently, computed retinal excitation energies can offer valuable insight into the validity of a given protein structural model. While agreement with experimental absorption energies does not confirm a structure, significant discrepancies may suggest that the model is inaccurate. Given that the structures of bR and ChR2 are well established; and Chrimson shares many structural motifs with ChR2, while exhibiting even more red-shifted absorption wrt bR, comparing these three systems provides meaningful insights to access the structural models developed in this work.[21,209,216,217]

When applying quantum chemical (QC) methods within QM/MM approaches, three primary sources of error must be considered: (i) QC methods often lack accuracy in absolute excitation energies. This is evident in the vacuum excitation energy, which typically exhibits a systematic error. However, this can be corrected by calibrating against accurate gas-phase reference values for the retinal chromophore. (ii) Accurately modeling the response to the electrostatic potential from the protein environment, the so-called opsin shift, is particularly challenging. The retinal chromophore is highly polarizable, and this poses a significant difficulty even for well-established quantum chemistry methods (see, e.g., Ref. [231, 244]). (iii) Finally, the effect of protein polarization and dispersion can contribute significantly to the excitation energy, as discussed in Refs.[218,219]. Despite these known limitations, we apply LC-TD-DFTB within an electrostatic embedding scheme to enable extensive conformational sampling. Due to the high flexibility of the active site, it is not feasible to identify few representative structures suitable for high-level quantum chemical calculations. As a result, our simulations are not intended to yield highly accurate absolute excitation energies. Instead, we focus on comparing relative trends across the three proteins. While LC-TD-DFTB systematically underestimates excitation energy shifts, it reliably captures qualitative trends.

Chrimson absorbs at 590 nm (2.10 eV) under neutral pH conditions, bR at 568 nm (2.18 eV) and ChR2 at 473 nm (2.62 eV).[209] A previous benchmark of LC-TD-DFTB has shown that opsin shifts are typically underestimated by about 50%. For example, the computed shift between bR and ChR2 is expected to be around 0.2 eV, compared to the experimental value of 0.4 eV. It is worth noting that most DFT methods, such as CAM-B3LYP or B3LYP, exhibit similar errors. Among tested functionals, ωB97 performed best, but still underestimates the shift by approximately 25%.[231] Based on previous benchmarking, LC-TD-DFTB is expected to yield a blue shift of 0.2 eV for ChR2 and a slight red shift of about 0.04 eV for Chrimson. The results summarized in Table 7.4 are consistent with these expectations. The red shift to ChR2 is underestimated, amounting to roughly half of the experimental value. In comparison, calculations using the ωB97X functional reproduce about 75% of the experimental bR-ChR2 shift, while the ChR2-Chrimson shift reaches 60-70% of the measured value, depending on the

Table 7.4: Comparison of absorption maxima of retinal proteins and retinal in vacuo. Excitation energies (EE) in eV of QM/MM sampled structures. The calculations are performed at the OM2/MRCI and LC-TD-DFTB/MM level of theory. WT_X denotes the shifts of Chrimson Models IV and V wrt retinal in vacuo.

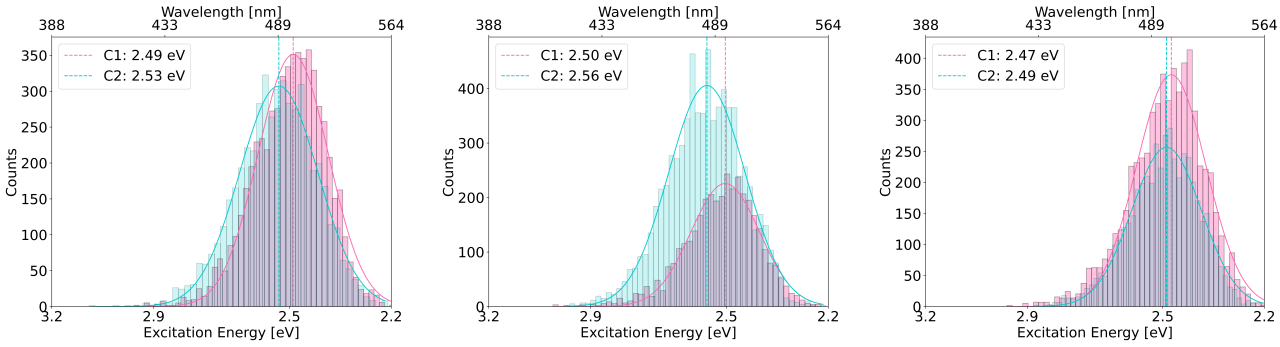
		bR	**ChR2**	**Chrimson**	
				WT_{IV}	**WT_V**
exp[209]		2.18	2.62	2.10	
$\Delta_{\text{protein-retinal}}$*		0.18	0.62	0.10	
OM2/MRCI**	vac	2.25	2.29	2.26	2.26
	QM/MM	2.63	3.30	2.71	2.76
	$\Delta_{\text{QM/MM-vac}}$	0.38	1.01	0.45	0.50
	$\Delta_{\text{Chrimson-bR}}$			0.08	0.13
	$\Delta_{\text{Chrimson-ChR2}}$			0.59	0.54
LC-TD-DFTB	vac	2.30	2.34	2.31	2.32
	QM/MM	2.52	2.75	2.51	2.54
	$\Delta_{\text{QM/MM-vac}}$	0.22	0.41	0.20	0.22
	$\Delta_{\text{Chrimson-bR}}$			-0.01	0.02
	$\Delta_{\text{Chrimson-ChR2}}$			-0.24	-0.21
ωB97X/def2-TZVP***	vac	2.60	2.67	2.65	2.63
	QM/MM	2.87	3.21	2.85	2.90
	$\Delta_{\text{QM/MM-vac}}$	0.27	0.54	0.20	0.27
	$\Delta_{\text{Chrimson-bR}}$			-0.02	0.03
	$\Delta_{\text{Chrimson-ChR2}}$			-0.36	-0.31

* Absorption maximum of all-trans retinal with protonated Schiff base in vacuo is about 2.00 eV.[254]
** Only protein and QM waters are considered as point charges, water and popc membrane are excluded.
*** Sampling with 250 randomly chosen snapshots for each system due to the higher computational cost.

model. For the bR-Chrimson shift, both methods, LC-TD-DFTB and ω97X, underestimate the red shift, with model V even yielding a slight blue shift.

We also report opsin shifts and results from OM2/MRCI calculations, which due to technical limitations were performed without including the membrane. OM2/MRCI yields a qualitatively similar picture, though it overestimates the opsin shift, as expected. Nevertheless, since the overall trends are consistent with those from the other methods, these additional calculations support the validity of our modeling strategy.

The excitation energies are similar for both models IV and V, and comparison with bR and ChR2 gives qualitatively correct spectral shifts. The comparison with bR reveals very similar results for the Chrimson models. Running our simulations and subsequent excitation energy calculations in triplicate, we observed a standard deviation of approximately 0.02 eV. When this variability is considered alongside the known underestimation of spectral shifts by the LC-TD-DFTB method, the computed red shift of 0.01 eV appears slightly smaller than expected based on the combined systematic and statistical uncertainties. Nonetheless, the discrepancy is minor, and we consider both models IV and V suitable candidates for further investigation in mutation studies.



(a) Model IV **(b)** Model V **(c)** S169A

Figure 7.10: Histograms of the excitation energies in eV according to the structural motif in the active site. Condition for the configuration is the distance between the counterions E165 and D295. The mean values are indicated on the top right and the colors indicate the configuration of the active site.

The excitation energies were correlated with the corresponding structural motifs in the active site to identify structural features responsible for the red-shifted absorption in Chrimson. Histograms of the excitation energies (see Fig. 7.10) are categorized by the structural motif present in the active site, as defined by the distance between the counterion oxygen atoms. Conformation C1, characterized by a counterion oxygen-oxygen distance of less than 3 Å, clearly correlates with lower excitation energies and thus contributes to the red shift. In model IV, conformation C1 occurs more frequently, resulting in a slightly more pronounced red shift compared to model V.

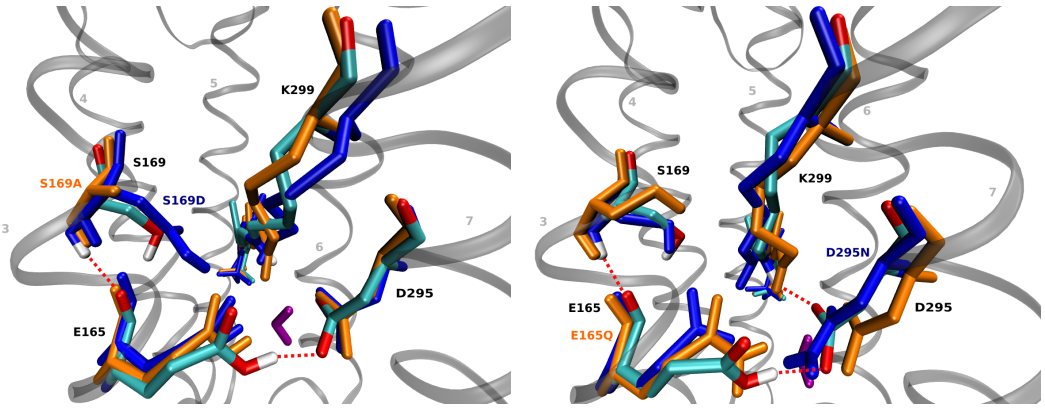
7.3.3 Chrimson Mutants

Based on the wt-Chrimson structure discussed above, we investigated mutants of residue S169 and the counterions, key positions where mutations disrupt the counterion network and shift hydrogen-bonding patterns. Moreover, these changes lead to more red-shifted variants, enhancing their potential for optogenetic applications. The primary goal of the mutant study is validating the computational model from the previous section by comparing excitation energies of the mutants to those of wt-Chrimson.

Chrimson mutants were constructed using glutamate protonation states defined in model V, which was favored due to minor structural differences, although models IV and V exhibit similar trends in their excitation energies.

7.3.3.1 Structural Analysis of S169 Mutants

Residue S169 forms a direct hydrogen bond with E165 and has been shown to influence the polarity of the active site, which is one of the main factors responsible for color tuning in rhodopsins.[19] The S169A mutation leads to a red-shifted absorption,[209] and in the newly discovered infrared absorbing rhodopsin proteins NeoR[211] and BestR,[210] the homologous residue of S169 is replaced by an aspartate. To investigate these red shifts, we therefore studied the two mutations S169A and S169D, the latter with protonated Asp. The active sites of both S169A(D) mutants compared to wt-Chrimson are shown in Fig. 7.11(a).



(a) S169X **(b)** Counterion mutants

Figure 7.11: The central gate of wt-Chrimson model V (cyan) and of various mutants. Possible hydrogen bonds are indicated by red dotted lines. Water is shown in purple. (a) S169X mutants. (b) Counterion mutants.

During the QM/MM MD simulations of S169A, both counterion conformations (C1 and C2, see above) were observed as in Chrimson WT. The distances of the counterions to the $RSBH^+$ are reduced in the S169A mutant (≈ 3 Å) and the occurrence of hydrogen bonding of the $RSBH^+$ is significantly increased (see Table 7.5). In contrast to wt-Chrimson where two hydrogen bonds can be formed between E165 and S169, the mutant features only one hydrogen bond, which is formed between the backbones of E165 and A169. The reason is that alanine is smaller than serine and lacks a proton donor for the additional hydrogen bond. As a result, conformation C1 occurs more frequently than C2 (see Table 7.2 in section 7.3.2). The second hydrogen bond between S169 and E165 promotes conformation C2 as discussed in section 7.3.2.1.

In contrast, the mutation S169D reduces the distance between the oxygens of E165 and D169 to 2 Å, the residues being permanently hydrogen bonded. This leads to a restriction of the torsion angle of E165 to values between -60 and -100°, which favors the C2 conformation being present in ≈ 94 % of the sampled conformations (see Table 7.2 in section 7.3.2). Hydrogen bonding of $RSBH^+$ occurs with a frequency similar to that in the wt structures, however, favoring the hydrogen bonds with the water molecule (compare Table 7.5).

Table 7.5: Hydrogen bonds between the $RSBH^+$ as acceptor and the counterions (E165 and D295) or nearby water as donor. Criteria to be met for hydrogen bonding: (i) $\text{angle}_{+\text{HN}\cdots\text{O}} \leq 30°$ and (ii) $\text{distance}_{\text{N}\cdots\text{O}} \leq 3.5$ Å. All results in % of the occurrences along the trajectory

	S169A	S169D	E165Q	D295N
$RSBH^+\cdots$E165	32.1	12.9	40.2	92.6
$RSBH^+\cdots$D295	34.1	17.4	21.4	0.1
$RSBH^+\cdots$SQM	42.0	32.1	2.5	0.7
total	86.0	61.8	63.9	92.8

7.3.3.2 Structural Analysis of the Counterion Mutants

Mutation of one of the counterions results in a blue-shifted absorption.[209] To study this effect, both counterions have been mutated to their corresponding amide form while adjusting the charge of the remaining counterion. For E165Q, counterion D295 remains deprotonated, while for D295N the counterion E165 is deprotonated to stabilize $RSBH^+$.

Fig. 7.11(b) shows the active site of counterion mutants compared to Chrimson WT. In both mutants, direct hydrogen bonding between E165(Q) and D295(N) (conformation C1) is rarely observed and the water molecule moves out of the active site, out of reach for hydrogen bonding to $RSBH^+$ (see Table 7.5). The $RSBH^+$ then forms hydrogen bonds primarily with E165(Q) and to a lesser extent with D295(N). In D295N only one hydrogen bond is formed between $RSBH^+$ and E165. Mutating either of the two counterions to larger residues results in a more compact active site, which prevents water from forming hydrogen bonds directly with the $RSBH^+$.

7.3.3.3 Excitation Energies of the Chrimson Mutants

Experimental studies show that S169A is red-shifted compared to the wt and absorbs at 608 nm (2.04 eV), while mutation of either one of the counterions, E165Q and D295N, leads to a blue-shifted spectrum (2.25 eV and 2.30 eV, respectively).[209] Table 7.6 shows the results of the excited state calculations of the mutants compared to wt-Chrimson. In agreement with experiment, our calculations show a red-shifted absorption of S169A, while for S169D only a very small shift in the range of the standard deviation is found, as discussed in section 7.3.2.2. As expected, the experimental shift of E165Q and D295N is qualitatively reproduced but underestimated, as discussed above. The deviation is part due to the limitations of the applied excited states method, but the structural models of the mutations may also contribute.

Table 7.6: Comparison of absorption maxima of Chrimson mutants: Excitation Energies (EE) in eV of QM/MM sampled structures. Calculations are performed at the LC-TD-DFTB level of theory.

	wt_V	S169A	S169D	E165Q	D295N
exp[209]	2.10	2.04	-	2.25	2.30
QM/MM	2.54	2.48	2.53	2.57	2.61
$\Delta_{\text{wt-mutant}}$		-0.06	-0.01	0.03	0.07

7.3.3.4 Influence of Structural Motifs on the Absorption of S169A

Fig. 7.10(c) shows histograms of excitation energies for different active site conformations in the S169A mutant, distinguished by the distance between the counterions. As in wt-Chrimson, conformation C1 is red-shifted relative to C2 and occurs more frequently (see Table 7.2 in section 7.3.2)). This shift can be attributed to the loss of one hydrogen bond: unlike serine in wt-Chrimson, alanine at position 169 lacks an additional proton donor, allowing only a single hydrogen bond between the backbones of E165 and A169. Nevertheless, both conformations in the mutant are more red-shifted than in wt-Chrimson, due to reduced polarity near the $RSBH^+$.

In summary, the study of the Chrimson mutants provides an interesting insight into altered hydrogen patterns reflected in the experimental absorption shifts. Our analysis of the mutants validates the computational model proposed in section 7.3.2.1 and confirms the findings regarding the red shift in Chrimson from section 7.3.2.2: the presence of the direct hydrogen bond between the counterions shifts the absorption spectra towards lower excitation energies.

7.4 Conclusions

We applied a multiscale approach based on DFTB/MD simulations to investigate the active site of the red-light-sensitive channelrhodopsin Chrimson, which is of particular significance for optogenetic applications. Our study provides new insights into the molecular dynamics at the nanosecond timescale.

Several studies have investigated the molecular origin of the red-shift in Chrimson[209,260,261], identifying three key factors: (i) the protonation of one of the counterion residues destabilizing the $RSBH^+$ in the ground state; (ii) a biased distribution of polar residues near the β-ionone ring, similar to bR; and (iii) a rigid retinal chromophore stabilized by bulky side chains. The mutant S169A, reported in Ref. [209], exhibits an even greater red-shift, attributed to a further reduction in active site polarity.

Our calculations successfully reproduce the characteristic red-shift of Chrimson relative to other ChRs. Moreover, we validate the reliability of our computational model by simulating several Chrimson mutants, recovering the experimentally observed spectral shifts. Consistent with experimental data, we confirm that E165 serves as the protonated counterion. As discussed above, the structure of the retinal binding pocket, particularly the residues in the active site, is a key determinant of color tuning in Chrimson. To investigate this, we examined the structural dynamics, which cannot be resolved by static techniques such as X-ray diffraction. As a result, we identified a multitude of active site conformations that deviate from the the deposited crystal structure. The active site of Chrimson exhibits structural heterogeneity comparable to that observed in ChR2,[216] including water movement within the active site and pronounced fluctuations in the distances between the counterions and the retinal Schiff base.

Interestingly, contrary to the crystallographic structure,[209] both counterions can approach the $RSBH^+$ within ≤ 3 Å, and D295 exhibits a higher frequency of hydrogen bonding with the $RSBH^+$ than E165. Two dominant structural motifs emerge along the trajectory: either the counterions are in close proximity ($E165_{OE}$-$D295_{OD}$ distance ≤ 3 Å), or they are farther apart, bridged by a hydrogen-bonded water molecule. Notably, at the counterion-counterion distance of approximately 2.8 Å, strong hydrogen bonding is observed, although no proton sharing occurs. Our analysis reveals a correlation between structural motifs and optical properties: the conformation characterized by a reduced distance between the counterions correlates with longer-wavelength absorption and is more prevalent in the red-shifted mutant S169A. This enhanced occurrence is linked to the S169A substitution: the replacement of serin by alanine results in the loss of an additional hydrogen bond to E165 that restricts the torsional flexibility of E165, promoting the red-shifted conformation. These findings represent a significant step towards understanding structure-function relationships in color tuning, providing a mechanistic basis for rational design of optogenetic tools with tailored spectral properties.

Overall, our QM/MM simulations have provided a rationalization for the crystal structure as an average of the conformations found in our simulations. These conformations interconvert on the nanosecond timescale, likely contributing to the intermediate active site distances in the crystal structure, which appear slightly too long for strong hydrogen bonds.[291]

Our work addresses several challenges in the multi-scale modeling of ChRs, some of which are discussed in a recent review on photoreceptor modeling.[23] One key issue is the treatment of multiple protonation states. In principle, future simulations may be based on constant pH molecular dynamics (CpHMD), as suggested in the review. Although approaches such as CpHMD followed by QM/MM simulations currently involve significantly higher computational costs, ongoing advances in hardware performance may make them increasingly feasible and help to reduce the uncertainty associated with using several models with different protonation states.

A major limitation, however, lies in the inability of current force fields to reliably model extended networks of strong hydrogen bonds, as demonstrated in this and previous studies.[216,217] While the use of polarized charges,[293] offers some improvements, their application remains problematic in highly flexible systems such as channel proteins. Consequently, we employed constrained MD simulations during equilibration and used large QM regions to accurately refine the active site.

Machine learning-based models may eventually help overcome this limitation; however, accurate equivariant neural network potentials currently exhibit computational demands comparable to the DFTB3 method used in this work. For such highly flexible active sites, there appears to be no viable substitute for fast QM or ML-based models combined with extensive sampling. The commonly used approach of relying on QM/MM-optimized structures, which is sufficient for many photoreceptors, is clearly inadequate in this case. An additional challenge arises from the complex electronic structure of retinal, which poses significant difficulties for accurate modeling.[231,244]

All TD-DFT methods, including those applying range-separated functionals, have been shown to have large errors considering shifts in excitation energies. Even the generally very well-performing ωB97 functional underestimates color shifts by approximately 25 %, while other commonly used functionals such as B3LPY or CAM-B3LYP can exhibit errors of up to 50%.[231]

Future work in our group aims to extend LC-TD-DFTB by implementing more accurate functionals, such as ωB97, into the DFTB framework. However, due to inherent limitations of the method, the potential for improvement may be limited. As an alternative, the development and application of machine learning models trained on post-Hartree Fock methods seems to be a promising approach.

8 Summary and Outlook

This PhD thesis illustrates how computational chemistry tools can be designed to explore the role of nearby residues and solvent exposure on two fundamental biochemical processes: proton-coupled electron transfer (PCET) and light absorption in retinal proteins. In both cases, multi-scale simulation protocols were developed and tested on minimal systems before application to full proteins. These protocols are computationally affordable, due to the application of semi-empirical density functional tight binding (DFTB) in combination with enhanced sampling methods or machine learning (ML). At the same time, successful validation and comparison to theoretical and experimental results demonstrate that accuracy does not suffer, which renders the workflows optimal for simulations of challenging biosystems.

Chapter 3 presents a systematic investigation of PCET in biomimetic peptides with different reaction partners (radical tyrosine with histidine, tryptophan, or tyrosine), geometries, solvent exposures, and protein embeddings. The multi-scale protocol combines hybrid quantum mechanics/molecular mechanics (QM/MM) simulations with metadynamics, using semi-empirical DFTB to access nanosecond timescales. Complete conformational sampling is achieved by biasing the proton transfer collective variable (CV). The free energy surfaces (FESs) are reconstructed by reweighting the electron transfer CV and provide both reaction barriers and mechanistic insights. It is shown that residue orientation and environmental factors influence the PCET mechanism.

In chapter 4, to transfer the tested protocol to full proteins, two proteins were selected, the class II photolyase from archaea *Methanosarcina mazei* (MmCPDII) and animal-like cryptochrome from green alga *Chlamydomonas reinhardtii* (CraCRY). In these proteins, PCET occurs between a radical tryptophan cation ($WH^{\bullet+}$) and a tyrosine residue (Y). Due to the rapidity and enhanced sampling of the protocol, multiple possible pathways were tested. In MmCPDII, the FES yielded low reaction barriers in the case of proton transfer (PT) from $WH^{\bullet+}$ to histidine with simultaneous electron transfer (ET) from Y to $W^{\bullet}$. Another possibility, in which aspartate acts as the proton acceptor instead of histidine, is excluded due to high barriers and unstable products. In CraCRY, a mechanism in which water-mediated deprotonation of $WH^{\bullet+}$ into bulk solvent precedes direct PCET between $W^{\bullet}$ and Y, appears very likely. Even though the two-dimensional FESs reveals key reactant/product states and reaction barriers for the concerted PCET mechanism, the sampling of intermediates of sequential PCET remains incomplete.

In chapter 5, a workflow is introduced that enables simultaneous biasing of PT and ET CVs using a DFTB implementation with coupled-perturbed (CP) equations. This should accelerate the convergence of the two-dimensional FESs. The CP equations are solved for all quantum mechanical (QM) atoms and selected molecular mechanics (MM) atoms. Tests varying the number of MM atoms in a cutoff sphere show negligible differences in free energy profiles but significant performance gains with smaller cutoffs. However, convergence issues arise when biasing only the ET coordinate, and barriers are overestimated for biomimetic peptides.

Looking ahead, several strategies could still enhance the performance of these protocols, such as refining restraints and metadynamics parameters for PCET metadynamics with dual PT/ET bias. Identification of components of the CP-equations, which contribute negligibly at longer distances and annulation of these, could lower computational demand. Alternative ET descriptors beyond Mulliken charges should be explored to improve the description of electronic states.

Chapter 6 investigates color tuning in retinal proteins using a machine learning (ML) approach. Instead of performing excited-state calculations directly on the proteins, all-trans retinal and 11-cis retinal are first simulated in solvents with varied ion configurations to mimic protein binding pockets. Excitation energies and oscillator strengths, computed with OM2/MRCI and LC-TD-DFTB, serve as training data for neural networks. Retinal proteins are then simulated only at ground-state level

to extract input features for the trained ML models, which predicted their spectroscopic properties. This strategy enables qualitative ranking of proteins by their experimental color tuning while restricting computationally expensive excited-state calculations to the training set.

In chapter 7, a structural and spectroscopic study of the channelrhodopsin Chrimson examines how active-site configurations influence absorption. QM/MM simulations reveal multiple protonated retinal Schiff base ($RSBH^+$)-counterion arrangements, including one featuring a hydrogen-bonded counterion pair. This configuration correlates to Chrimson's red-shifted absorption and appears more frequently in the S169A mutant. In this mutant, it can be traced back to enhanced torsional flexibility of the counterion E165 due to the loss of a stabilizing hydrogen bond. For retinal spectroscopy, incorporating higher-level post-Hartree Fock (HF) methods could increase accuracy, though at a higher computational cost. In addition, extending simulations to excited-state dynamics could provide valuable insight into the primary photoprocesses of retinal proteins.

Across both research themes, the results demonstrate that electrostatics and hydrogen bonding networks in the local environment are central to biomolecular reactivity. The inherent dynamics of biological systems, in which the environment modulates the biochemical reaction, is challenging to capture in simulations. This thesis presents several tailor-made protocols that address and overcome these challenges. Hybrid QM/MM methods provide a practical framework for capturing these effects. The use of DFTB as QM method, in combination with either metadynamics to enhance sampling or ML algorithms to speed up calculations, allows to take into account a wide conformational landscape. With that, valuable insights into PCET mechanisms and thermodynamics in two proteins were obtained. Additionally the absorption capabilities of the chromophore retinal in various protein-like environments were assessed. Nonetheless, there are several open issues remaining, including the necessity to improve the description of the ET CV, the consideration of nonadiabatic reactions including electronic excited states in PCET, as well as the incorporation of more accurate theoretical methods in the ML approach. Yet, the protocols presented here demonstrate great potential and subsequent improvements could effectively address these issues, yielding an even better description of reactivity in proteins.

List of Publications

Journal Articles

[1] Katharina Spies et al. “Structural and environmental effects on the mechanism of biological proton-coupled electron transfer using DFTB/MM metadynamics”. In: *Physical Chemistry Chemical Physics* 27 (2025). Publisher: The Royal Society of Chemistry, pp. 15544–15556. DOI: 10.1039/D5CP01359C.

[2] Katharina Spies, Beatrix M. Bold, and Marcus Elstner. “Complex active site structures influence absorption spectrum of Chrimson wild type and mutants”. In: *Phys. Chem. Chem. Phys.* 27 (June 2025), pp. 13360–13370. ISSN: 1463-9084. DOI: 10.1039/D5CP00762C.

Conference Contributions

- Modena, Italy; July 3-5 2023; **Principles of Light-Induced Charge Transfer for Optogenetics**; Talk: "Active site structure of Chrimson wild type and mutants and their absorption spectra"
- Lyon, France; August 27-31 2023; **20th Congress of the European Society for Photobiology**; Talk: "Active Site Structure and Absorption Spectrum of the Channelrhodopsin Chrimson: Wild Type and Mutants"
- Hünfeld, Germany; April 19-20 2024; **Computer Simulation and Theory of Macromolecules**; Talk: "Proton Coupled Electron Transfer in Biomimetic Peptides"

Bibliography

[1] JoAnne Stubbe and Wilfred A. Van Der Donk. “Protein Radicals in Enzyme Catalysis”. en. In: *Chemical Reviews* 98.2 (Apr. 1998), pp. 705–762. ISSN: 0009-2665, 1520-6890. DOI: 10.1021/cr9400875.

[2] David R. Weinberg et al. “Proton-Coupled Electron Transfer”. In: *Chemical Reviews* 112.7 (July 2012), pp. 4016–4093. ISSN: 0009-2665. DOI: 10.1021/cr200177j.

[3] Robin S. Sibert, Mira Josowicz, and Bridgette A. Barry. “Control of proton and electron transfer in de novo designed, biomimetic β hairpins”. eng. In: *ACS chemical biology* 5.12 (Dec. 2010), pp. 1157–1168. ISSN: 1554-8937. DOI: 10.1021/cb100138m.

[4] Sharon Hammes-Schiffer. “Exploring proton-coupled electron transfer at multiple scales”. en. In: *Nature Computational Science* 3.4 (Apr. 2023), pp. 291–300. ISSN: 2662-8457. DOI: 10.1038/s43588-023-00422-5.

[5] Astrid Nilsen-Moe et al. “Switching the proton-coupled electron transfer mechanism for non-canonical tyrosine residues in a de novo protein”. en. In: *Chemical Science* 15.11 (Jan. 2024). Publisher: The Royal Society of Chemistry, pp. 3957–3970. ISSN: 2041-6539. DOI: 10.1039/D3SC05450K.

[6] Clorice R. Reinhardt et al. “Computing Proton-Coupled Redox Potentials of Fluorotyrosines in a Protein Environment”. en. In: *The Journal of Physical Chemistry B* 125.1 (Jan. 2021), pp. 128–136. ISSN: 1520-6106, 1520-5207. DOI: 10.1021/acs.jpcb.0c09974.

[7] José-Alain Sahel et al. “Partial recovery of visual function in a blind patient after optogenetic therapy”. en. In: *Nature Medicine* 27.7 (July 2021), pp. 1223–1229. ISSN: 1078-8956, 1546-170X. DOI: 10.1038/s41591-021-01351-4.

[8] Anding Bi et al. “Ectopic Expression of a Microbial-Type Rhodopsin Restores Visual Responses in Mice with Photoreceptor Degeneration”. English. In: *Neuron* 50.1 (Apr. 2006). Publisher: Elsevier, pp. 23–33. ISSN: 0896-6273. DOI: 10.1016/j.neuron.2006.02.026.

[9] Juliette E. McGregor et al. “Optogenetic restoration of retinal ganglion cell activity in the living primate”. en. In: *Nature Communications* 11.1 (Apr. 2020), p. 1703. ISSN: 2041-1723. DOI: 10.1038/s41467-020-15317-6.

[10] T. Shimano et al. “Assessment of the AAV-mediated expression of channelrhodopsin-2 and halorhodopsin in brainstem neurons mediating auditory signaling”. In: *Brain Research*. Optogenetics and Pharmacogenetics in Neuronal Function and Dysfunction 1511 (May 2013), pp. 138–152. ISSN: 0006-8993. DOI: 10.1016/j.brainres.2012.10.030.

[11] Maria Zerche, Tobias Moser, and Thomas Mager. “Identification and optimization of channelrhodopsin variants for optogenetic hearing restoration”. en. In: *Laryngo-Rhino-Otologie*. Vol. 101. ISSN: 0935-8943, 1438-8685 Issue: S 2 Journal Abbreviation: Laryngorhinootologie. Georg Thieme Verlag, May 2022. DOI: 10.1055/s-0042-1746832.

[12] Burak Bali et al. “Analyzing efficacy, stability, and safety of AAV-mediated optogenetic hearing restoration in mice”. en. In: *Life Science Alliance* 5.8 (Aug. 2022), e202101338. ISSN: 2575-1077. DOI: 10.26508/lsa.202101338.

[13] Maria Zerche et al. “Channelrhodopsin fluorescent tag replacement for clinical translation of optogenetic hearing restoration”. en. In: *Molecular Therapy - Methods & Clinical Development* 29 (June 2023), pp. 202–212. ISSN: 23290501. DOI: 10.1016/j.omtm.2023.03.009.

[14] Viviana Gradinaru et al. “Optical Deconstruction of Parkinsonian Neural Circuitry”. In: *Science* 324.5925 (Apr. 2009). Publisher: American Association for the Advancement of Science, pp. 354–359. DOI: 10.1126/science.1167093.

[15] E. Kip et al. “Patterned Stimulation of the Chrimson Opsin in Glutamatergic Motor Thalamus Neurons Improves Forelimb Akinesia in Parkinsonian Rats”. en. In: *Neuroscience* 507 (Dec. 2022), pp. 64–78. ISSN: 03064522. DOI: 10.1016/j.neuroscience.2022.10.026.

[16] Clorice R. Reinhardt et al. “Glutamate Mediates Proton-Coupled Electron Transfer Between Tyrosines 730 and 731 in *Escherichia coli* Ribonucleotide Reductase”. en. In: *Journal of the American Chemical Society* 143.16 (Apr. 2021), pp. 6054–6059. ISSN: 0002-7863, 1520-5126. DOI: 10.1021/jacs.1c02152.

[17] Jiayun Zhong, Alexander V. Soudackov, and Sharon Hammes-Schiffer. “Probing Nonadiabaticity of Proton-Coupled Electron Transfer in Ribonucleotide Reductase”. In: *The Journal of Physical Chemistry Letters* 15.6 (Feb. 2024), pp. 1686–1693. DOI: 10.1021/acs.jpclett.3c03552.

[18] Bruce W. Berry, Melissa C. Martínez-Rivera, and Cecilia Tommos. “Reversible voltammograms and a Pourbaix diagram for a protein tyrosine radical”. In: *Proceedings of the National Academy of Sciences* 109.25 (June 2012), pp. 9739–9743. DOI: 10.1073/pnas.1112057109.

[19] Oliver P. Ernst et al. “Microbial and Animal Rhodopsins: Structures, Functions, and Molecular Mechanisms”. en. In: *Chemical Reviews* 114.1 (Jan. 2014), pp. 126–163. ISSN: 0009-2665, 1520-6890. DOI: 10.1021/cr4003769.

[20] Franziska Schneider, Christiane Grimm, and Peter Hegemann. “Biophysics of Channelrhodopsin”. en. In: *Annual Review of Biophysics* 44.1 (June 2015), pp. 167–186. ISSN: 1936-122X, 1936-1238. DOI: 10.1146/annurev-biophys-060414-034014.

[21] Michael Hoffmann et al. “Color Tuning in Rhodopsins: The Mechanism for the Spectral Shift between Bacteriorhodopsin and Sensory Rhodopsin II”. en. In: *Journal of the American Chemical Society* 128.33 (Aug. 2006), pp. 10808–10818. ISSN: 0002-7863, 1520-5126. DOI: 10.1021/ja062082i.

[22] Ville R. I. Kaila. “Resolving Chemical Dynamics in Biological Energy Conversion: Long-Range Proton-Coupled Electron Transfer in Respiratory Complex I”. en. In: *Accounts of Chemical Research* 54.24 (Dec. 2021), pp. 4462–4473. ISSN: 0001-4842, 1520-4898. DOI: 10.1021/acs.accounts.1c00524.

[23] Maria-Andrea Mroginski et al. “Frontiers in Multiscale Modeling of Photoreceptor Proteins”. en. In: *Photochemistry and Photobiology* 97.2 (Mar. 2021), pp. 243–269. ISSN: 0031-8655, 1751-1097. DOI: 10.1111/php.13372.

[24] Wolfram Koch and Max C. Holthausen. *A Chemist’s Guide to Density Functional Theory*. en. 1st ed. Wiley, July 2001. ISBN: 978-3-527-30372-4 978-3-527-60004-5. DOI: 10.1002/3527600043.

[25] Frank Jensen. *Introduction to Computational Chemistry*. en. Google-Books-ID: RDIG48UcZfYC. Wiley, Jan. 2007. ISBN: 978-0-470-05804-6.

[26] Michael Levitt and Arieh Warshel. “Computer simulation of protein folding”. en. In: *Nature* 253.5494 (Feb. 1975), pp. 694–698. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/253694a0.

[27] Michael Levitt. “A simplified representation of protein conformations for rapid simulation of protein folding”. en. In: *Journal of Molecular Biology* 104.1 (June 1976), pp. 59–107. ISSN: 00222836. DOI: 10.1016/0022-2836(76)90004-8.

[28] W. G. Noid. “Perspective: Advances, Challenges, and Insight for Predictive Coarse-Grained Models”. en. In: *The Journal of Physical Chemistry B* 127.19 (May 2023), pp. 4174–4207. ISSN: 1520-6106, 1520-5207. DOI: 10.1021/acs.jpcb.2c08731.

[29] John A. Keith et al. “Combining Machine Learning and Computational Chemistry for Predictive Insights Into Chemical Systems”. en. In: *Chemical Reviews* 121.16 (Aug. 2021), pp. 9816–9872. ISSN: 0009-2665, 1520-6890. DOI: 10.1021/acs.chemrev.1c00107.

[30] A. Warshel and M. Levitt. “Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme”. en. In: *Journal of Molecular Biology* 103.2 (May 1976), pp. 227–249. ISSN: 00222836. DOI: 10.1016/0022-2836(76)90311-9.

[31] Gerrit Groenhof. “Introduction to QM/MM Simulations”. en. In: *Biomolecular Simulations*. Ed. by Luca Monticelli and Emppu Salonen. Vol. 924. Series Title: Methods in Molecular Biology. Totowa, NJ: Humana Press, 2013, pp. 43–66. ISBN: 978-1-62703-016-8 978-1-62703-017-5. DOI: 10.1007/978-1-62703-017-5_3.

[32] Loup Verlet. “Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules”. en. In: *Physical Review* 159.1 (July 1967), pp. 98–103. ISSN: 0031-899X. DOI: 10.1103/PhysRev.159.98.

[33] A. Iserles. “Generalized Leapfrog Methods”. en. In: *IMA Journal of Numerical Analysis* 6.4 (1986), pp. 381–392. ISSN: 0272-4979, 1464-3642. DOI: 10.1093/imanum/6.4.381.

[34] William L. Jorgensen et al. “Comparison of simple potential functions for simulating liquid water”. In: *The Journal of Chemical Physics* 79.2 (July 1983), pp. 926–935. ISSN: 0021-9606. DOI: 10.1063/1.445869.

[35] H. J. C. Berendsen et al. “Interaction Models for Water in Relation to Protein Hydration”. en. In: *Intermolecular Forces: Proceedings of the Fourteenth Jerusalem Symposium on Quantum Chemistry and Biochemistry Held in Jerusalem, Israel, April 13–16, 1981*. Ed. by Bernard Pullman. Dordrecht: Springer Netherlands, 1981, pp. 331–342. ISBN: 978-94-015-7658-1. DOI: 10.1007/978-94-015-7658-1_21.

[36] Saeed Izadi, Ramu Anandakrishnan, and Alexey V. Onufriev. “Building Water Models: A Different Approach”. In: *The Journal of Physical Chemistry Letters* 5.21 (Nov. 2014). Publisher: American Chemical Society, pp. 3863–3871. DOI: 10.1021/jz501780a.

[37] Berk Hess et al. “LINCS: A linear constraint solver for molecular simulations”. en. In: *Journal of Computational Chemistry* 18.12 (Sept. 1997), pp. 1463–1472. ISSN: 0192-8651, 1096-987X. DOI: 10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H.

[38] Jean-Paul Ryckaert, Giovanni Ciccotti, and Herman J. C Berendsen. “Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of *n*-alkanes”. In: *Journal of Computational Physics* 23.3 (Mar. 1977), pp. 327–341. ISSN: 0021-9991. DOI: 10.1016/0021-9991(77)90098-5.

[39] Hans C Andersen. “Rattle: A “velocity” version of the shake algorithm for molecular dynamics calculations”. en. In: *Journal of Computational Physics* 52.1 (Oct. 1983), pp. 24–34. ISSN: 00219991. DOI: 10.1016/0021-9991(83)90014-1.

[40] G.M. Torrie and J.P. Valleau. “Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling”. en. In: *Journal of Computational Physics* 23.2 (Feb. 1977), pp. 187–199. ISSN: 00219991. DOI: 10.1016/0021-9991(77)90121-8.

[41] *Umbrella Sampling - an overview | ScienceDirect Topics*.

[42] Alessandro Laio and Michele Parrinello. “Escaping free-energy minima”. en. In: *Proceedings of the National Academy of Sciences* 99.20 (Oct. 2002), pp. 12562–12566. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.202427399.

[43] Alessandro Barducci, Giovanni Bussi, and Michele Parrinello. “Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method”. en. In: *Physical Review Letters* 100.2 (Jan. 2008), p. 020603. ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.100.020603.

[44] Pratyush Tiwary and Michele Parrinello. “A Time-Independent Free Energy Estimator for Metadynamics”. en. In: *The Journal of Physical Chemistry B* 119.3 (Jan. 2015), pp. 736–742. ISSN: 1520-6106, 1520-5207. DOI: 10.1021/jp504920s.

[45] M. Bonomi, A. Barducci, and M. Parrinello. “Reconstructing the equilibrium Boltzmann distribution from well-tempered metadynamics”. en. In: *Journal of Computational Chemistry* 30.11 (Aug. 2009), pp. 1615–1621. ISSN: 0192-8651, 1096-987X. DOI: 10.1002/jcc.21305.

[46] Omar Valsson, Pratyush Tiwary, and Michele Parrinello. “Enhancing Important Fluctuations: Rare Events and Metadynamics from a Conceptual Viewpoint”. en. In: *Annual Review of Physical Chemistry* 67.1 (May 2016), pp. 159–184. ISSN: 0066-426X, 1545-1593. DOI: 10.1146/annurev-physchem-040215-112229.

[47] Giovanni Bussi and Gareth A. Tribello. “Analyzing and Biasing Simulations with PLUMED”. en. In: *Biomolecular Simulations*. Ed. by Massimiliano Bonomi and Carlo Camilloni. Vol. 2022. Series Title: Methods in Molecular Biology. New York, NY: Springer New York, 2019, pp. 529–578. ISBN: 978-1-4939-9607-0 978-1-4939-9608-7. DOI: 10.1007/978-1-4939-9608-7_21.

[48] Donald B. Rubin. “The Bayesian Bootstrap”. en. In: *The Annals of Statistics* 9.1 (Jan. 1981), pp. 130–134. ISSN: 0090-5364. DOI: 10.1214/aos/1176345338.

[49] Werner Kutzelnigg. *Einführung in die theoretische Chemie*. en. Kompaktausg. Weinheim, Germany: Wiley-VCH, 2002. ISBN: 978-3-527-30609-1.

[50] Hermann Haken and Hans C. Wolf. *Molekülphysik und Quantenchemie von Hermann Haken | ISBN 978-3-540-30314-5 | Fachbuch online kaufen - Lehmanns.de*. de. 2006.

[51] Attila Szabo and Neil S. Ostlund. *Modern Quantum Chemistry von Attila Szabo | ISBN 978-0-486-69186-2 | Fachbuch online kaufen - Lehmanns.de*. de. 1989.

[52] Yoshihiro Tawada et al. “A long-range-corrected time-dependent density functional theory”. en. In: *The Journal of Chemical Physics* 120.18 (May 2004), pp. 8425–8433. ISSN: 0021-9606, 1089-7690. DOI: 10.1063/1.1688752.

[53] Walter Thiel. “Semiempirical quantum–chemical methods”. en. In: *WIREs Computational Molecular Science* 4.2 (Mar. 2014), pp. 145–157. ISSN: 1759-0876, 1759-0884. DOI: 10.1002/wcms.1161.

[54] Pavlo O. Dral et al. “Semiempirical Quantum-Chemical Orthogonalization-Corrected Methods: Theory, Implementation, and Parameters”. en. In: *Journal of Chemical Theory and Computation* 12.3 (Mar. 2016), pp. 1082–1096. ISSN: 1549-9618, 1549-9626. DOI: 10.1021/acs.jctc.5b01046.

[55] M. Elstner et al. “Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties”. en. In: *Physical Review B* 58.11 (Sept. 1998), pp. 7260–7268. ISSN: 0163-1829, 1095-3795. DOI: 10.1103/PhysRevB.58.7260.

[56] M. Elstner et al. “A Self-Consistent Charge Density-Functional Based Tight-Binding Scheme for Large Biomolecules”. en. In: *physica status solidi (b)* 217.1 (Jan. 2000), pp. 357–376. ISSN: 0370-1972, 1521-3951. DOI: 10.1002/(SICI)1521-3951(200001)217:1<357::AID-PSSB357>3.0.CO;2-J.

[57] M. Elstner. “The SCC-DFTB method and its application to biological systems”. en. In: *Theoretical Chemistry Accounts* 116.1-3 (July 2006), pp. 316–325. ISSN: 1432-881X, 1432-2234. DOI: 10.1007/s00214-005-0066-0.

[58] Michael Gaus, Qiang Cui, and Marcus Elstner. “DFTB3: Extension of the Self-Consistent-Charge Density-Functional Tight-Binding Method (SCC-DFTB)”. en. In: *Journal of Chemical Theory and Computation* 7.4 (Apr. 2011), pp. 931–948. ISSN: 1549-9618, 1549-9626. DOI: 10.1021/ct100684s.

[59] Michael Gaus, Albrecht Goez, and Marcus Elstner. “Parametrization and Benchmark of DFTB3 for Organic Molecules”. en. In: *Journal of Chemical Theory and Computation* 9.1 (Jan. 2013), pp. 338–354. ISSN: 1549-9618, 1549-9626. DOI: 10.1021/ct300849w.

[60] Denis Maag et al. “Mechanism of proton-coupled electron transfer described with QM/MM implementation of coupled-perturbed density-functional tight-binding”. en. In: *The Journal of Chemical Physics* (Mar. 2023), p. 5.0137122. ISSN: 0021-9606, 1089-7690. DOI: 10.1063/5.0137122.

[61] Thomas A. Niehaus and Fabio Della Sala. “Range separated functionals in the density functional based tight-binding method: Formalism”. en. In: *physica status solidi (b)* 249.2 (Feb. 2012), pp. 237–244. ISSN: 03701972. DOI: 10.1002/pssb.201100694.

[62] V. Lutsker, B. Aradi, and T. A. Niehaus. “Implementation and benchmark of a long-range corrected functional in the density functional based tight-binding method”. en. In: *The Journal of Chemical Physics* 143.18 (Nov. 2015), p. 184107. ISSN: 0021-9606, 1089-7690. DOI: 10.1063/1.4935095.

[63] Julian J. Kranz et al. “Time-Dependent Extension of the Long-Range Corrected Density Functional Based Tight-Binding Method”. en. In: *Journal of Chemical Theory and Computation* 13.4 (Apr. 2017), pp. 1737–1747. ISSN: 1549-9618, 1549-9626. DOI: 10.1021/acs.jctc.6b01243.

[64] Henryk A. Witek, Stephan Irle, and Keiji Morokuma. “Analytical second-order geometrical derivatives of energy for the self-consistent-charge density-functional tight-binding method”. en. In: *The Journal of Chemical Physics* 121.11 (Sept. 2004), pp. 5163–5170. ISSN: 0021-9606, 1089-7690. DOI: 10.1063/1.1775786.

[65] B. Hourahine et al. “DFTB+, a software package for efficient approximate density functional theory based atomistic simulations”. en. In: *The Journal of Chemical Physics* 152.12 (Mar. 2020), p. 124101. ISSN: 0021-9606, 1089-7690. DOI: 10.1063/1.5143190.

[66] Natacha Gillet, Marcus Elstner, and Tomáš Kubař. “Coupled-perturbed DFTB-QM/MM metadynamics: Application to proton-coupled electron transfer”. en. In: *The Journal of Chemical Physics* 149.7 (Aug. 2018), p. 072328. ISSN: 0021-9606, 1089-7690. DOI: 10.1063/1.5027100.

[67] T. Kubař, M. Elstner, and Q. Cui. “Hybrid Quantum Mechanical/Molecular Mechanical Methods For Studying Energy Transduction in Biomolecular Machines”. en. In: *Annual Review of Biophysics* 52.1 (May 2023), pp. 525–551. ISSN: 1936-122X, 1936-1238. DOI: 10.1146/annurev-biophys-111622-091140.

[68] Boyi Zhou, Yanzi Zhou, and Daiqian Xie. “Accelerated Quantum Mechanics/Molecular Mechanics Simulations via Neural Networks Incorporated with Mechanical Embedding Scheme”. In: *Journal of Chemical Theory and Computation* 19.4 (Feb. 2023). Publisher: American Chemical Society, pp. 1157–1169. ISSN: 1549-9618. DOI: 10.1021/acs.jctc.2c01131.

[69] John Jumper et al. “Highly accurate protein structure prediction with AlphaFold”. en. In: *Nature* 596.7873 (Aug. 2021), pp. 583–589. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/s41586-021-03819-2.

[70] Mihaly Varadi et al. “AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models”. en. In: *Nucleic Acids Research* 50.D1 (Jan. 2022), pp. D439–D444. ISSN: 0305-1048, 1362-4962. DOI: 10.1093/nar/gkab1061.

[71] *Press release: The Nobel Prize in Chemistry 2024*. en-US.

[72] Matthias Rupp. “Machine learning for quantum mechanics in a nutshell”. en. In: *International Journal of Quantum Chemistry* 115.16 (Aug. 2015), pp. 1058–1073. ISSN: 0020-7608, 1097-461X. DOI: 10.1002/qua.24954.

[73] Jörg Behler and Michele Parrinello. “Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces”. en. In: *Physical Review Letters* 98.14 (Apr. 2007), p. 146401. ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.98.146401.

[74] Kristof T. Schütt et al., eds. *Machine Learning Meets Quantum Physics*. en. Vol. 968. Lecture Notes in Physics. Cham: Springer International Publishing, 2020. ISBN: 978-3-030-40244-0 978-3-030-40245-7. DOI: 10.1007/978-3-030-40245-7.

[75] Izaak Neutelings. *Neural networks – TikZ.net*. en-US. Apr. 2024.

[76] Sharon Hammes-Schiffer. “Proton-coupled electron transfer: classification scheme and guide to theoretical methods”. en. In: *Energy & Environmental Science* 5.7 (2012), p. 7696. ISSN: 1754-5692, 1754-5706. DOI: 10.1039/c2ee03361e.

[77] R. A. Marcus. “Chemical and Electrochemical Electron-Transfer Theory”. en. In: *Annual Review of Physical Chemistry* 15.Volume 15, (Oct. 1964). Publisher: Annual Reviews, pp. 155–196. ISSN: 0066-426X, 1545-1593. DOI: 10.1146/annurev.pc.15.100164.001103.

[78] Sharon Hammes-Schiffer. “Theoretical Perspectives on Proton-Coupled Electron Transfer Reactions”. en. In: *Accounts of Chemical Research* 34.4 (Apr. 2001), pp. 273–281. ISSN: 0001-4842, 1520-4898. DOI: 10.1021/ar9901117.

[79] JoAnne Stubbe et al. “Radical Initiation in the Class I Ribonucleotide Reductase: Long-Range Proton-Coupled Electron Transfer?” In: *Chemical Reviews* 103.6 (June 2003), pp. 2167–2202. ISSN: 0009-2665. DOI: 10.1021/cr020421u.

[80] Arieh Warshel. “Computer Simulations of Enzyme Catalysis: Methods, Progress, and Insights”. en. In: *Annual Review of Biophysics and Biomolecular Structure* 32.1 (June 2003), pp. 425–443. ISSN: 1056-8700, 1545-4266. DOI: 10.1146/annurev.biophys.32.110601.141807.

[81] P. H. König et al. “Toward Theoretical Analyis of Long-Range Proton Transfer Kinetics in Biomolecular Pumps”. en. In: *The Journal of Physical Chemistry A* 110.2 (Jan. 2006), pp. 548–563. ISSN: 1089-5639, 1520-5215. DOI: 10.1021/jp052328q.

[82] Denis Maag et al. “O to bR transition in bacteriorhodopsin occurs through a proton hole mechanism”. en. In: *Proceedings of the National Academy of Sciences* 118.39 (Sept. 2021), e2024803118. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.2024803118.

[83] Sharon Hammes-Schiffer. “Proton-Coupled Electron Transfer: Moving Together and Charging Forward”. en. In: *Journal of the American Chemical Society* 137.28 (July 2015), pp. 8860–8871. ISSN: 0002-7863, 1520-5126. DOI: 10.1021/jacs.5b04087.

[84] Sharon Hammes-Schiffer and Alexander V. Soudackov. “Proton-Coupled Electron Transfer in Solution, Proteins, and Electrochemistry”. In: *The Journal of Physical Chemistry B* 112.45 (Nov. 2008), pp. 14108–14123. ISSN: 1520-6106. DOI: 10.1021/jp805876e.

[85] Joseph R. Lakowicz and Barry R. Masters. “Principles of Fluorescence Spectroscopy, Third Edition”. en. In: *Journal of Biomedical Optics* 13.2 (2008), p. 029901. ISSN: 10833668. DOI: 10.1117/1.2904580.

[86] Maurizio Persico and Giovanni Granucci. *Photochemistry*. en. Theoretical Chemistry and Computational Modelling. Cham: Springer International Publishing, 2018. ISBN: 978-3-319-89971-8 978-3-319-89972-5. DOI: 10.1007/978-3-319-89972-5.

[87] Benjamin G. Levine and Todd J. Martínez. “Isomerization Through Conical Intersections”. en. In: *Annual Review of Physical Chemistry* 58.1 (May 2007), pp. 613–634. ISSN: 0066-426X, 1545-1593. DOI: 10.1146/annurev.physchem.57.032905.104612.

[88] Frank Neese. “A spectroscopy oriented configuration interaction procedure”. en. In: *The Journal of Chemical Physics* 119.18 (Nov. 2003), pp. 9428–9443. ISSN: 0021-9606, 1089-7690. DOI: 10.1063/1.1615956.

[89] Axel Koslowski, Michael E. Beck, and Walter Thiel. “Implementation of a general multireference configuration interaction procedure with analytic gradients in a semiempirical context using the graphical unitary group approach”. en. In: *Journal of Computational Chemistry* 24.6 (Apr. 2003), pp. 714–726. ISSN: 0192-8651, 1096-987X. DOI: 10.1002/jcc.10210.

[90] Filipp Furche and Reinhart Ahlrichs. “Adiabatic time-dependent density functional methods for excited state properties”. en. In: *The Journal of Chemical Physics* 117.16 (Oct. 2002), pp. 7433–7447. ISSN: 0021-9606, 1089-7690. DOI: 10.1063/1.1508368.

[91] M.E. Casida and M. Huix-Rotllant. “Progress in Time-Dependent Density-Functional Theory”. en. In: *Annual Review of Physical Chemistry* 63.1 (May 2012), pp. 287–323. ISSN: 0066-426X, 1545-1593. DOI: 10.1146/annurev-physchem-032511-143803.

[92] Kieron Burke and Lucas O. Wagner. “DFT in a nutshell”. en. In: *International Journal of Quantum Chemistry* 113.2 (Jan. 2013), pp. 96–101. ISSN: 0020-7608, 1097-461X. DOI: 10.1002/qua.24259.

[93] *Brief Introduction to TDDFT*.

[94] T. A. Niehaus et al. “Tight-binding approach to time-dependent density-functional response theory”. en. In: *Physical Review B* 63.8 (Feb. 2001), p. 085108. ISSN: 0163-1829, 1095-3795. DOI: 10.1103/PhysRevB.63.085108.

[95] T.A. Niehaus. “Approximate time-dependent density functional theory”. en. In: *Journal of Molecular Structure: THEOCHEM* 914.1-3 (Nov. 2009), pp. 38–49. ISSN: 01661280. DOI: 10.1016/j.theochem.2009.04.034.

[96] Tim J. Zuehlsdorff et al. “Optical spectra in the condensed phase: Capturing anharmonic and vibronic features using dynamic and static approaches”. en. In: *The Journal of Chemical Physics* 151.7 (Aug. 2019), p. 074111. ISSN: 0021-9606, 1089-7690. DOI: 10.1063/1.5114818.

[97] Katharina Spies et al. “Structural and environmental effects on the mechanism of biological proton-coupled electron transfer using DFTB/MM metadynamics”. In: *Physical Chemistry Chemical Physics* 27 (2025). Publisher: The Royal Society of Chemistry, pp. 15544–15556. DOI: 10.1039/D5CP01359C.

[98] Ulla Uhlin and Hans Eklund. “Structure of ribonucleotide reductase protein R”. en. In: 370 (1994).

[99] Robin Sibert et al. “Proton-Coupled Electron Transfer in a Biomimetic Peptide as a Model of Enzyme Regulatory Mechanisms”. In: *Journal of the American Chemical Society* 129.14 (Apr. 2007), pp. 4393–4400. ISSN: 0002-7863. DOI: 10.1021/ja068805f.

[100] Bridgette A. Barry. “Reaction dynamics and proton coupled electron transfer: Studies of tyrosine-based charge transfer in natural and biomimetic systems”. en. In: *Biochimica et Biophysica Acta (BBA) - Bioenergetics* 1847.1 (Jan. 2015), pp. 46–54. ISSN: 00052728. DOI: 10.1016/j.bbabio.2014.09.003.

[101] Cynthia V. Pagba et al. “Proton-Coupled Electron Transfer and a Tyrosine–Histidine Pair in a Photosystem II-Inspired β-Hairpin Maquette: Kinetics on the Picosecond Time Scale”. In: *The Journal of Physical Chemistry B* 120.7 (Feb. 2016), pp. 1259–1272. ISSN: 1520-6106. DOI: 10.1021/acs.jpcb.6b00560.

[102] Tyler G. McCaslin et al. “Structure and Function of Tryptophan–Tyrosine Dyads in Biomimetic β Hairpins”. en. In: *The Journal of Physical Chemistry B* 123.13 (Apr. 2019), pp. 2780–2791. ISSN: 1520-6106, 1520-5207. DOI: 10.1021/acs.jpcb.8b12452.

[103] Cecilia Tommos et al. “De Novo Proteins as Models of Radical Enzymes”. en. In: *Biochemistry* 38.29 (July 1999), pp. 9495–9507. ISSN: 0006-2960, 1520-4995. DOI: 10.1021/bi990609g.

[104] Anthony. Harriman. “Further comments on the redox potentials of tryptophan and tyrosine”. en. In: *The Journal of Physical Chemistry* 91.24 (Nov. 1987), pp. 6102–6104. ISSN: 0022-3654, 1541-5740. DOI: 10.1021/j100308a011.

[105] William T. Dixon and David Murphy. “Determination of the acidity constants of some phenol radical cations by means of electron spin resonance”. en. In: *Journal of the Chemical Society, Faraday Transactions 2* 72 (1976), p. 1221. ISSN: 0300-9238. DOI: 10.1039/f29767201221.

[106] Sabine Oldemeyer et al. “Essential Role of an Unusually Long-lived Tyrosyl Radical in the Response to Red Light of the Animal-like Cryptochrome aCRY”. en. In: *Journal of Biological Chemistry* 291.27 (July 2016), pp. 14062–14071. ISSN: 00219258. DOI: 10.1074/jbc.M116.726976.

[107] Kanchana R. Ravichandran et al. “Formal Reduction Potentials of Difluorotyrosine and Trifluorotyrosine Protein Residues: Defining the Thermodynamics of Multistep Radical Transfer”. In: *Journal of the American Chemical Society* 139.8 (Mar. 2017), pp. 2994–3004. ISSN: 0002-7863. DOI: 10.1021/jacs.6b11011.

[108] Astrid Nilsen-Moe et al. “Proton-Coupled Electron Transfer from Tyrosine in the Interior of a de novo Protein: Mechanisms and Primary Proton Acceptor”. In: *Journal of the American Chemical Society* 142.26 (July 2020), pp. 11550–11559. ISSN: 0002-7863. DOI: 10.1021/jacs.0c04655.

[109] Starla D. Glover et al. “Photochemical Tyrosine Oxidation in the Structurally Well-Defined α3Y Protein: Proton-Coupled Electron Transfer and a Long-Lived Tyrosine Radical”. In: *Journal of the American Chemical Society* 136.40 (Oct. 2014), pp. 14039–14051. ISSN: 0002-7863. DOI: 10.1021/ja503348d.

[110] Xiaohua Chen et al. “Proton-Regulated Electron Transfers from Tyrosine to Tryptophan in Proteins: Through-Bond Mechanism versus Long-Range Hopping Mechanism”. en. In: *The Journal of Physical Chemistry B* 113.52 (Dec. 2009), pp. 16681–16688. ISSN: 1520-6106, 1520-5207. DOI: 10.1021/jp9077689.

[111] James M. Mayer et al. “Proton-Coupled Electron Transfer versus Hydrogen Atom Transfer in Benzyl/Toluene, Methoxyl/Methanol, and Phenoxyl/Phenol Self-Exchange Reactions”. en. In: *Journal of the American Chemical Society* 124.37 (Sept. 2002), pp. 11142–11147. ISSN: 0002-7863, 1520-5126. DOI: 10.1021/ja012732c.

[112] Ville R. I. Kaila and Gerhard Hummer. “Energetics of Direct and Water-Mediated Proton-Coupled Electron Transfer”. en. In: *Journal of the American Chemical Society* 133.47 (Nov. 2011), pp. 19040–19043. ISSN: 0002-7863, 1520-5126. DOI: 10.1021/ja2082262.

[113] Alessandro Barducci, Massimiliano Bonomi, and Michele Parrinello. “Metadynamics”. en. In: *WIREs Computational Molecular Science* 1.5 (Sept. 2011), pp. 826–843. ISSN: 1759-0876, 1759-0884. DOI: 10.1002/wcms.31.

[114] Giovanni Bussi and Alessandro Laio. “Using metadynamics to explore complex free-energy landscapes”. en. In: *Nature Reviews Physics* 2.4 (Mar. 2020), pp. 200–212. ISSN: 2522-5820. DOI: 10.1038/s42254-020-0153-0.

[115] Timo M. Schäfer and Giovanni Settanni. “Data Reweighting in Metadynamics Simulations”. en. In: *Journal of Chemical Theory and Computation* 16.4 (Apr. 2020), pp. 2042–2052. ISSN: 1549-9618, 1549-9626. DOI: 10.1021/acs.jctc.9b00867.

[116] F. Giberti et al. “Iterative Unbiasing of Quasi-Equilibrium Sampling”. en. In: *Journal of Chemical Theory and Computation* 16.1 (Jan. 2020), pp. 100–107. ISSN: 1549-9618, 1549-9626. DOI: 10.1021/acs.jctc.9b00907.

[117] Gareth A. Tribello et al. “PLUMED 2: New feathers for an old bird”. In: *Computer Physics Communications* 185.2 (Feb. 2014), pp. 604–613. ISSN: 0010-4655. DOI: 10.1016/j.cpc.2013.09.018.

[118] Nicolas Mandard et al. “The solution structure of gomesin, an antimicrobial cysteine-rich peptide from the spider”. en. In: *European Journal of Biochemistry* 269.4 (2002). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1046/j.0014-2956.2002.02760.x, pp. 1190–1198. ISSN: 1432-1033. DOI: 10.1046/j.0014-2956.2002.02760.x.

[119] D.A. Case, H.M. Aktulga, and P.A. Kollman. *Amber 2023*. University of California, San Francisco, 2023.

[120] Chuan Tian et al. “ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution”. en. In: *Journal of Chemical Theory and Computation* 16.1 (Jan. 2020), pp. 528–552. ISSN: 1549-9618, 1549-9626. DOI: 10.1021/acs.jctc.9b00591.

[121] Junmei Wang et al. “Automatic atom type and bond type perception in molecular mechanical calculations”. eng. In: *Journal of Molecular Graphics & Modelling* 25.2 (Oct. 2006), pp. 247–260. ISSN: 1093-3263. DOI: 10.1016/j.jmgm.2005.12.005.

[122] Junmei Wang et al. “Development and testing of a general amber force field”. en. In: *Journal of Computational Chemistry* 25.9 (2004). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jcc.20035, pp. 1157–1174. ISSN: 1096-987X. DOI: 10.1002/jcc.20035.

[123] M. J. Frisch et al. *Gaussian 09*. Gaussian Inc. Wallingford CT 2009.

[124] Sander Pronk et al. “GROMACS 4.5”. In: *Bioinformatics* 29.7 (Apr. 2013), pp. 845–854. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btt055.

[125] Mark James Abraham et al. “GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers”. In: *SoftwareX* 1-2 (Sept. 2015), pp. 19–25. ISSN: 2352-7110. DOI: 10.1016/j.softx.2015.06.001.

[126] Christof Köhler et al. “Treatment of Collinear and Noncollinear Electron Spin within an Approximate Density Functional Based Method”. en. In: *The Journal of Physical Chemistry A* 111.26 (July 2007), pp. 5622–5629. ISSN: 1089-5639, 1520-5215. DOI: 10.1021/jp068802p.

[127] Tomáš Kubař. *tomaskubar/dftbplus*. original-date: 2020-02-20T09:43:17Z. Oct. 2022.

[128] Tomáš Kubař. *tomaskubar/gromacs-dftbplus*. original-date: 2019-04-24T16:05:16Z. Feb. 2025.

[129] Van Quan Vuong et al. “Parametrization and Benchmark of Long-Range Corrected Dftb2 for Organic Molecules”. en. In: *Journal of Chemical Theory and Computation* 14.1 (Jan. 2018), pp. 115–125. ISSN: 1549-9618, 1549-9626. DOI: 10.1021/acs.jctc.7b00947.

[130] Yan Zhao and Donald G. Truhlar. “The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals”. en. In: *Theoretical Chemistry Accounts* 120.1 (May 2008), pp. 215–241. ISSN: 1432-2234. DOI: 10.1007/s00214-007-0310-x.

[131] A. D. McLean and G. S. Chandler. “Contracted Gaussian basis sets for molecular calculations. I. Second row atoms, Z=11–18”. In: *The Journal of Chemical Physics* 72.10 (May 1980), pp. 5639–5648. ISSN: 0021-9606. DOI: 10.1063/1.438980.

[132] R. Krishnan et al. “Self-consistent molecular orbital methods. XX. A basis set for correlated wave functions”. In: *The Journal of Chemical Physics* 72.1 (Jan. 1980), pp. 650–654. ISSN: 0021-9606. DOI: 10.1063/1.438955.

[133] Jeng-Da Chai and Martin Head-Gordon. “Systematic optimization of long-range corrected hybrid density functionals”. In: *The Journal of Chemical Physics* 128.8 (Feb. 2008), p. 084106. ISSN: 0021-9606. DOI: 10.1063/1.2834918.

[134] Guido Van Rossum and Fred L. Drake. *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace, 2009. ISBN: 1-4414-1269-7.

[135] J. D. Hunter. “Matplotlib: A 2D graphics environment”. In: *Computing in Science & Engineering* 9.3 (2007). Publisher: IEEE COMPUTER SOC, pp. 90–95. DOI: 10.1109/MCSE.2007.55.

[136] Charles R. Harris et al. “Array programming with NumPy”. In: *Nature* 585 (2020), pp. 357–362. DOI: 10.1038/s41586-020-2649-2.

[137] Wes McKinney et al. “Data structures for statistical computing in python”. In: *Proceedings of the 9th Python in Science Conference*. Vol. 445. Austin, TX, 2010, pp. 51–56.

[138] Michael L. Waskom. “seaborn: statistical data visualization”. In: *Journal of Open Source Software* 6.60 (2021). Publisher: The Open Journal, p. 3021. DOI: 10.21105/joss.03021.

[139] Cynthia V. Pagba et al. “A tyrosine–tryptophan dyad and radical-based charge transfer in a ribonucleotide reductase-inspired maquette”. en. In: *Nature Communications* 6.1 (Dec. 2015), p. 10010. ISSN: 2041-1723. DOI: 10.1038/ncomms10010.

[140] Idelisa Pujols-Ayala and Bridgette A. Barry. “Tyrosyl radicals in Photosystem II”. en. In: *Biochimica et Biophysica Acta (BBA) - Bioenergetics* 1655 (Apr. 2004), pp. 205–216. ISSN: 00052728. DOI: 10.1016/j.bbabio.2003.07.010.

[141] Harry B. Gray and Jay R. Winkler. “Hole hopping through tyrosine/tryptophan chains protects proteins from oxidative damage”. en. In: *Proceedings of the National Academy of Sciences* 112.35 (Sept. 2015), pp. 10920–10925. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1512704112.

[142] Fabien Lacombat et al. “Ultrafast Oxidation of a Tyrosine by Proton-Coupled Electron Transfer Promotes Light Activation of an Animal-like Cryptochrome”. In: *Journal of the American Chemical Society* 141.34 (Aug. 2019), pp. 13394–13409. ISSN: 0002-7863. DOI: 10.1021/jacs.9b03680.

[143] Pavel Müller et al. “Sub-nanosecond tryptophan radical deprotonation mediated by a protein-bound water cluster in class II DNA photolyases”. en. In: *Chemical Science* 9.5 (2018), pp. 1200–1212. ISSN: 2041-6520, 2041-6539. DOI: 10.1039/C7SC03969G.

[144] Clorice R. Reinhardt et al. “Conformational Motions and Water Networks at the ∝/β Interface in E. coli Ribonucleotide Reductase”. In: *Journal of the American Chemical Society* 142.32 (Aug. 2020), pp. 13768–13778. ISSN: 0002-7863. DOI: 10.1021/jacs.0c04325.

[145] Cynthia V. Pagba et al. “Proton-Coupled Electron Transfer in Tyrosine and a β-Hairpin Maquette: Reaction Dynamics on the Picosecond Time Scale”. en. In: *The Journal of Physical Chemistry B* 119.6 (Feb. 2015), pp. 2726–2736. ISSN: 1520-6106, 1520-5207. DOI: 10.1021/jp510171z.

[146] Sebastian Salentin et al. “PLIP: fully automated protein–ligand interaction profiler”. In: *Nucleic Acids Research* 43.Web Server issue (July 2015), W443–W447. ISSN: 0305-1048. DOI: 10.1093/nar/gkv315.

[147] B. Hourahine et al. “Self-Interaction and Strong Correlation in DFTB”. en. In: *The Journal of Physical Chemistry A* 111.26 (July 2007), pp. 5671–5677. ISSN: 1089-5639, 1520-5215. DOI: 10.1021/jp070173b.

[148] Aron J. Cohen, Paula Mori-Sánchez, and Weitao Yang. “Insights into Current Limitations of Density Functional Theory”. In: *Science* 321.5890 (Aug. 2008). Publisher: American Association for the Advancement of Science, pp. 792–794. DOI: 10.1126/science.1158722.

[149] Mathias Rapacioli et al. “Modeling Charge Resonance in Cationic Molecular Clusters: Combining DFT-Tight Binding with Configuration Interaction”. en. In: *Journal of Chemical Theory and Computation* 7.1 (Jan. 2011), pp. 44–55. ISSN: 1549-9618, 1549-9626. DOI: 10.1021/ct100412f.

[150] Alexander V. Soudackov and Sharon Hammes-Schiffer. “Probing Nonadiabaticity in the Proton-Coupled Electron Transfer Reaction Catalyzed by Soybean Lipoxygenase”. en. In: *The Journal of Physical Chemistry Letters* 5.18 (Sept. 2014), pp. 3274–3278. ISSN: 1948-7185, 1948-7185. DOI: 10.1021/jz501655v.

[151] Mohanram Sivaraja et al. “Identification by ENDOR of Trp^{191} as the Free-Radical Site in Cytochrome c Peroxidase Compound ES”. en. In: *Science* 245.4919 (Aug. 1989), pp. 738–740. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.2549632.

[152] Fahmi Himo and Leif A. Eriksson. “Theoretical Study of Model Tryptophan Radicals and Radical Cations: Comparison with Experimental Data of DNA Photolyase, Cytochrome *c* Peroxidase, and Ribonucleotide Reductase”. en. In: *The Journal of Physical Chemistry B* 101.47 (Nov. 1997), pp. 9811–9819. ISSN: 1520-6106, 1520-5207. DOI: 10.1021/jp971041t.

[153] Andrii Piatkivskyi et al. “Structure and Reactivity of the Distonic and Aromatic Radical Cations of Tryptophan”. en. In: *Journal of the American Society for Mass Spectrometry* 24.4 (Apr. 2013), pp. 513–523. ISSN: 1044-0305. DOI: 10.1007/s13361-013-0594-0.

[154] Mark A. Miller, Lidia Vitello, and James E. Erman. “Regulation of Interprotein Electron Transfer By Trp 191 of Cytochrome c Peroxidase”. en. In: *Biochemistry* 34.37 (Sept. 1995), pp. 12048–12058. ISSN: 0006-2960, 1520-4995. DOI: 10.1021/bi00037a048.

[155] Peter Deppisch, Charlotte Helfrich-Förster, and Pingkalai R. Senthilan. “The Gain and Loss of Cryptochrome/Photolyase Family Members during Evolution”. en. In: *Genes* 13.9 (Sept. 2022), p. 1613. ISSN: 2073-4425. DOI: 10.3390/genes13091613.

[156] Aziz Sancar. “Structure and Function of DNA Photolyase and Cryptochrome Blue-Light Photoreceptors”. en. In: *Chemical Reviews* 103.6 (June 2003). Publisher: American Chemical Society (ACS), pp. 2203–2238. ISSN: 0009-2665, 1520-6890. DOI: 10.1021/cr0204348.

[157] Stefan Weber. “Light-driven enzymatic catalysis of DNA repair: a review of recent biophysical studies on photolyase”. en. In: *Biochimica et Biophysica Acta (BBA) - Bioenergetics* 1707.1 (Feb. 2005). Publisher: Elsevier BV, pp. 1–23. ISSN: 0005-2728. DOI: 10.1016/j.bbabio.2004.02.010.

[158] Dongping Zhong. “Electron Transfer Mechanisms of DNA Repair by Photolyase”. en. In: *Annual Review of Physical Chemistry* 66.1 (Apr. 2015). Publisher: Annual Reviews, pp. 691–715. ISSN: 0066-426X, 1545-1593. DOI: 10.1146/annurev-physchem-040513-103631.

[159] Meng Zhang, Lijuan Wang, and Dongping Zhong. “Photolyase: Dynamics and electron-transfer mechanisms of DNA repair”. en. In: *Archives of Biochemistry and Biophysics* 632 (Oct. 2017). Publisher: Elsevier BV, pp. 158–174. ISSN: 0003-9861. DOI: 10.1016/j.abb.2017.08.007.

[160] Chentao Lin and Dror Shalitin. “Cryptochrome Structure and Signal Transduction”. en. In: *Annual Review of Plant Biology* 54.1 (June 2003). Publisher: Annual Reviews, pp. 469–496. ISSN: 1543-5008, 1545-2123. DOI: 10.1146/annurev.arplant.54.110901.160901.

[161] Qin Wang and Chentao Lin. “Mechanisms of Cryptochrome-Mediated Photoresponses in Plants”. en. In: *Annual Review of Plant Biology* 71.1 (Apr. 2020). Publisher: Annual Reviews, pp. 103–129. ISSN: 1543-5008, 1545-2123. DOI: 10.1146/annurev-arplant-050718-100300.

[162] Sophie Franz et al. “Structure of the bifunctional cryptochrome aCRY from Chlamydomonas reinhardtii”. en. In: *Nucleic Acids Research* 46.15 (Sept. 2018), pp. 8010–8022. ISSN: 0305-1048, 1362-4962. DOI: 10.1093/nar/gky621.

[163] Nuri Ozturk et al. “Reaction mechanism of Drosophila cryptochrome”. en. In: *Proceedings of the National Academy of Sciences* 108.2 (Jan. 2011). Publisher: Proceedings of the National Academy of Sciences, pp. 516–521. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1017093108.

[164] Marion F. Haug et al. “Eumetazoan Cryptochrome Phylogeny and Evolution”. en. In: *Genome Biology and Evolution* 7.2 (Feb. 2015). Publisher: Oxford University Press (OUP), pp. 601–619. ISSN: 1759-6653. DOI: 10.1093/gbe/evv010.

[165] Thorsten Ritz, Salih Adem, and Klaus Schulten. “A Model for Photoreceptor-Based Magnetoreception in Birds”. en. In: *Biophysical Journal* 78.2 (Feb. 2000). Publisher: Elsevier BV, pp. 707–718. ISSN: 0006-3495. DOI: 10.1016/s0006-3495(00)76629-x.

[166] Miriam Liedvogel and Henrik Mouritsen. “Cryptochromes—a potential magnetoreceptor: what do we know and what do we want to know?” en. In: *Journal of The Royal Society Interface* 7.suppl_2 (Apr. 2010). Publisher: The Royal Society. ISSN: 1742-5689, 1742-5662. DOI: 10.1098/rsif.2009.0411.focus.

[167] Atticus Pinzon-Rodriguez, Staffan Bensch, and Rachel Muheim. “Expression patterns of cryptochrome genes in avian retina suggest involvement of Cry4 in light-dependent magnetoreception”. en. In: *Journal of The Royal Society Interface* 15.140 (Mar. 2018). Publisher: The Royal Society, p. 20180058. ISSN: 1742-5689, 1742-5662. DOI: 10.1098/rsif.2018.0058.

[168] Ilia A. Solov'yov, Tatiana Domratcheva, and Klaus Schulten. "Separation of photo-induced radical pair in cryptochrome to a functionally critical distance". en. In: *Scientific Reports* 4.1 (Jan. 2014). Publisher: Springer Science and Business Media LLC. ISSN: 2045-2322. DOI: 10.1038/srep03845.

[169] Gongyi Hong et al. "Electron transfer and spin dynamics of the radical-pair in the cryptochrome from Chlamydomonas reinhardtii by computational analysis". eng. In: *The Journal of Chemical Physics* 152.6 (Feb. 2020), p. 065101. ISSN: 1089-7690. DOI: 10.1063/1.5133019.

[170] Chentao Lin and Takeshi Todo. "The cryptochromes". en. In: *Genome Biology* 6.5 (2005). Publisher: Springer Science and Business Media LLC, p. 220. ISSN: 1465-6906. DOI: 10.1186/gb-2005-6-5-220.

[171] Inês Chaves et al. "The Cryptochromes: Blue Light Photoreceptors in Plants and Animals". en. In: *Annual Review of Plant Biology* 62.1 (June 2011). Publisher: Annual Reviews, pp. 335–364. ISSN: 1543-5008, 1545-2123. DOI: 10.1146/annurev-arplant-042110-103759.

[172] L Essen. "Photolyases and cryptochromes: common mechanisms of DNA repair and light-driven signaling?" en. In: *Current Opinion in Structural Biology* 16.1 (Feb. 2006). Publisher: Elsevier BV, pp. 51–59. ISSN: 0959-440X. DOI: 10.1016/j.sbi.2006.01.004.

[173] Corinne Aubert et al. "Intraprotein electron transfer between tyrosine and tryptophan in DNA photolyase from *Anacystis nidulans*". en. In: *Proceedings of the National Academy of Sciences* 96.10 (May 1999), pp. 5423–5427. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.96.10.5423.

[174] Margaret S. Cheung et al. "Pathways of Electron Transfer in Escherichia coli DNA Photolyase:Trp306 to FADH". en. In: *Biophysical Journal* 76.3 (Mar. 1999). Publisher: Elsevier BV, pp. 1241–1249. ISSN: 0006-3495. DOI: 10.1016/s0006-3495(99)77287-5.

[175] Martin Byrdin et al. "Dissection of the triple tryptophan electron transfer chain in *Escherichia coli* DNA photolyase: Trp382 is the primary donor in photoactivation". en. In: *Proceedings of the National Academy of Sciences* 100.15 (July 2003). Publisher: Proceedings of the National Academy of Sciences, pp. 8676–8681. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1531645100.

[176] Baldissera Giovani et al. "Light-induced electron transfer in a cryptochrome blue-light photoreceptor". en. In: *Nature Structural & Molecular Biology* 10.6 (June 2003), pp. 489–490. ISSN: 1545-9993, 1545-9985. DOI: 10.1038/nsb933.

[177] Goutham Kodali, Salim U. Siddiqui, and Robert J. Stanley. "Charge Redistribution in Oxidized and Semiquinone E. coli DNA Photolyase upon Photoexcitation: Stark Spectroscopy Reveals a Rationale for the Position of Trp382". en. In: *Journal of the American Chemical Society* 131.13 (Apr. 2009). Publisher: American Chemical Society (ACS), pp. 4795–4807. ISSN: 0002-7863, 1520-5126. DOI: 10.1021/ja809214r.

[178] Martin Byrdin et al. "Quantum Yield Measurements of Short-Lived Photoactivation Intermediates in DNA Photolyase: Toward a Detailed Understanding of the Triple Tryptophan Electron Transfer Chain". en. In: *The Journal of Physical Chemistry A* 114.9 (Mar. 2010). Publisher: American Chemical Society (ACS), pp. 3207–3214. ISSN: 1089-5639, 1520-5215. DOI: 10.1021/jp9093589.

[179] Daniel Nohr et al. "Extended Electron-Transfer in Animal Cryptochromes Mediated by a Tetrad of Aromatic Amino Acids". en. In: *Biophysical Journal* 111.2 (July 2016), pp. 301–311. ISSN: 00063495. DOI: 10.1016/j.bpj.2016.06.009.

[180] Brian D. Zoltowski et al. "Chemical and structural analysis of a photoactive vertebrate cryptochrome from pigeon". en. In: *Proceedings of the National Academy of Sciences* 116.39 (Sept. 2019). Publisher: Proceedings of the National Academy of Sciences, pp. 19449–19457. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1907875116.

[181] Zheyun Liu et al. "Electron Tunneling Pathways and Role of Adenine in Repair of Cyclobutane Pyrimidine Dimer by DNA Photolyase". en. In: *Journal of the American Chemical Society* 134.19 (May 2012). Publisher: American Chemical Society (ACS), pp. 8104–8114. ISSN: 0002-7863, 1520-5126. DOI: 10.1021/ja2105009.

[182] Zheyun Liu, Lijuan Wang, and Dongping Zhong. "Dynamics and mechanisms of DNA repair by photolyase". en. In: *Physical Chemistry Chemical Physics* 17.18 (2015). Publisher: Royal Society of Chemistry (RSC), pp. 11933–11949. ISSN: 1463-9076, 1463-9084. DOI: 10.1039/c4cp05286b.

[183] Meng Zhang, Lijuan Wang, and Dongping Zhong. "Photolyase: Dynamics and Mechanisms of Repair of Sun-Induced DNA Damage". en. In: *Photochemistry and Photobiology* 93.1 (Jan. 2017). Publisher: Wiley, pp. 78–92. ISSN: 0031-8655, 1751-1097. DOI: 10.1111/php.12695.

[184] Gesa Grüning et al. "The Effect of Spin Relaxation on Magnetic Compass Sensitivity in ErCry4a". en. In: *ChemPhysChem* 25.19 (Oct. 2024). Publisher: Wiley. ISSN: 1439-4235, 1439-7641. DOI: 10.1002/cphc.202400129.

[185] Jessica L. Ramsay et al. "Cryptochrome magnetoreception: Time course of photoactivation from non-equilibrium coarse-grained molecular dynamics". en. In: *Computational and Structural Biotechnology Journal* 26 (Dec. 2024). Publisher: Elsevier BV, pp. 58–69. ISSN: 2001-0370. DOI: 10.1016/j.csbj.2024.11.001.

[186] Fabian Schuhmann et al. "Structural Rearrangements of Pigeon Cryptochrome 4 Undergoing a Complete Redox Cycle". en. In: *The Journal of Physical Chemistry B* 128.16 (Apr. 2024). Publisher: American Chemical Society (ACS), pp. 3844–3855. ISSN: 1520-6106, 1520-5207. DOI: 10.1021/acs.jpcb.4c00424.

[187] Christopher B. Harrison, Lauren L. O'Nei, and Olaf Wiest. "Computational Studies of DNA Photolyase". en. In: *The Journal of Physical Chemistry A* 109.32 (Aug. 2005). Publisher: American Chemical Society (ACS), pp. 7001–7012. ISSN: 1089-5639, 1520-5215. DOI: 10.1021/jp051075y.

[188] Manuel Maestre-Reyna et al. "Serial crystallography captures dynamic control of sequential electron and proton transfer events in a flavoenzyme". en. In: *Nature Chemistry* 14.6 (June 2022), pp. 677–685. ISSN: 1755-4330, 1755-4349. DOI: 10.1038/s41557-022-00922-3.

[189] Benedikt Beel et al. "A Flavin Binding Cryptochrome Photoreceptor Responds to Both Blue and Red Light in *Chlamydomonas reinhardtii*". en. In: *The Plant Cell* 24.7 (July 2012), pp. 2992–3008. ISSN: 1040-4651, 1532-298X. DOI: 10.1105/tpc.112.098947.

[190] F. Kiefer et al. "The SWISS-MODEL Repository and associated resources". en. In: *Nucleic Acids Research* 37.Database (Jan. 2009), pp. D387–D392. ISSN: 0305-1048, 1362-4962. DOI: 10.1093/nar/gkn750.

[191] *H++ (web-based computational prediction of protonation states and pK of ionizable groups in macromolecules)*.

[192] J. C. Gordon et al. "H++: a server for estimating pKas and adding missing hydrogens to macromolecules". en. In: *Nucleic Acids Research* 33.Web Server (July 2005), W368–W371. ISSN: 0305-1048, 1362-4962. DOI: 10.1093/nar/gki464.

[193] R. Anandakrishnan, B. Aguilar, and A. V. Onufriev. "H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations". en. In: *Nucleic Acids Research* 40.W1 (July 2012), W537–W541. ISSN: 0305-1048, 1362-4962. DOI: 10.1093/nar/gks375.

[194] Jonathan Myers et al. "A simple clustering algorithm can be accurate enough for use in calculations of p K s in macromolecules". en. In: *Proteins: Structure, Function, and Bioinformatics* 63.4 (June 2006), pp. 928–938. ISSN: 0887-3585, 1097-0134. DOI: 10.1002/prot.20922.

[195] James A. Maier et al. "ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB". en. In: *Journal of Chemical Theory and Computation* 11.8 (Aug. 2015), pp. 3696–3713. ISSN: 1549-9618, 1549-9626. DOI: 10.1021/acs.jctc.5b00255.

[196] Mathew Chow, Clorice R. Reinhardt, and Sharon Hammes-Schiffer. "Nuclear Quantum Effects in Quantum Mechanical/Molecular Mechanical Free Energy Simulations of Ribonucleotide Reductase". en. In: *Journal of the American Chemical Society* (Nov. 2024), jacs.4c13955. ISSN: 0002-7863, 1520-5126. DOI: 10.1021/jacs.4c13955.

[197] Qiwen Zhu et al. "Proton-Coupled Electron Transfer upon Oxidation of Tyrosine in a De Novo Protein: Analysis of Proton Acceptor Candidates". en. In: *Biochemistry* 63.15 (Aug. 2024), pp. 1999–2008. ISSN: 0006-2960, 1520-4995. DOI: 10.1021/acs.biochem.4c00211.

[198] Daniel Riepl et al. "Long-range charge transfer mechanism of the III2IV2 mycobacterial supercomplex". en. In: *Nature Communications* 15.1 (June 2024), p. 5276. ISSN: 2041-1723. DOI: 10.1038/s41467-024-49628-9.

[199] Gyunghoon Kang et al. "Structure of a trapped radical transfer pathway within a ribonucleotide reductase holocomplex". In: *Science (New York, N.Y.)* 368.6489 (Apr. 2020), pp. 424–427. ISSN: 0036-8075. DOI: 10.1126/science.aba6794.

[200] Yanier Crespo et al. "Metadynamics convergence law in a multidimensional system". en. In: *Physical Review E* 81.5 (May 2010), p. 055701. ISSN: 1539-3755, 1550-2376. DOI: 10.1103/PhysRevE.81.055701.

[201] Fahimeh Baftizadeh et al. "Protein Folding and Ligand-Enzyme Binding from Bias-Exchange Metadynamics Simulations". en. In: *Current Physical Chemistry* 2.1 (Jan. 2012), pp. 79–91. ISSN: 18779468. DOI: 10.2174/1877946811202010079.

[202] Giovanni Bussi and Davide Branduardi. "Free-Energy Calculations with Metadynamics: Theory and Practice". en. In: *Reviews in Computational Chemistry*. Ed. by Abby L. Parrill and Kenny B. Lipkowitz. 1st ed. Wiley, Apr. 2015, pp. 1–49. ISBN: 978-1-118-40777-6 978-1-118-88988-6. DOI: 10.1002/9781118889886.ch1.

[203] Chenghan Li and Gregory A. Voth. "Using Constrained Density Functional Theory to Track Proton Transfers and to Sample Their Associated Free Energy Surface". en. In: *Journal of Chemical Theory and Computation* 17.9 (Sept. 2021), pp. 5759–5765. ISSN: 1549-9618, 1549-9626. DOI: 10.1021/acs.jctc.1c00609.

[204] Béryl Naëmi Greb. *Training of NN for the Prediction of 11-cis-Retinylidene Protein Absorption Spectra*. Bachelor Thesis. Karlsruhe Institute of Technology, July 2024.

[205] Lara Maria Nagel. *Trainingsdatengenerierung des neuronalen Netzes von 11-cis-Retinal basierend auf der Verwendung von QM/MM-Simulationen*. Bachelor Thesis. Karlsruhe Institute of Technology, Sept. 2024.

[206] Jemina Enkelmann. *Training Data Generation for Neural Networks describing all-trans Retinal using QM/MM simulations*. Bachelor Thesis. Karlsruhe Institute of Technology, Sept. 2024.

[207] Jingwei Jiang, Huxing Cui, and Kamal Rahmouni. "Optogenetics and pharmacogenetics: principles and applications". en. In: *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology* 313.6 (Dec. 2017), R633–R645. ISSN: 0363-6119, 1522-1490. DOI: 10.1152/ajpregu.00091.2017.

[208] Sonja Kleinlogel et al. "Emerging approaches for restoration of hearing and vision". en. In: *Physiological Reviews* (Mar. 2020), physrev.00035.2019. ISSN: 0031-9333, 1522-1210. DOI: 10.1152/physrev.00035.2019.

[209] Kazumasa Oda et al. "Crystal structure of the red light-activated channelrhodopsin Chrimson". en. In: *Nature Communications* 9.1 (Sept. 2018). Number: 1 Publisher: Nature Publishing Group, p. 3949. ISSN: 2041-1723. DOI: 10.1038/s41467-018-06421-9.

[210] Andrey Rozenberg et al. "Rhodopsin-bestrophin fusion proteins from unicellular algae form gigantic pentameric ion channels". en. In: *Nature Structural & Molecular Biology* 29.6 (June 2022), pp. 592–603. ISSN: 1545-9993, 1545-9985. DOI: 10.1038/s41594-022-00783-x.

[211] Matthias Broser et al. "NeoR, a near-infrared absorbing rhodopsin". en. In: *Nature Communications* 11.1 (Nov. 2020), p. 5682. ISSN: 2041-1723. DOI: 10.1038/s41467-020-19375-8.

[212] Ramkumar Rajamani, Yen-Lin Lin, and Jiali Gao. "The opsin shift and mechanism of spectral tuning in rhodopsin". en. In: *Journal of Computational Chemistry* 32.5 (Apr. 2011), pp. 854–865. ISSN: 0192-8651, 1096-987X. DOI: 10.1002/jcc.21663.

[213] Jan S. Frähmcke et al. "The Protonation State of Glu181 in Rhodopsin Revisited: Interpretation of Experimental Data on the Basis of QM/MM Calculations". en. In: *The Journal of Physical Chemistry B* 114.34 (Sept. 2010), pp. 11338–11352. ISSN: 1520-6106, 1520-5207. DOI: 10.1021/jp104537w.

[214] Jan S. Frähmcke, Marius Wanko, and Marcus Elstner. "Building a Model of the Blue Cone Pigment Based on the Wild Type Rhodopsin Structure with QM/MM Methods". en. In: *The Journal of Physical Chemistry B* 116.10 (Mar. 2012), pp. 3313–3321. ISSN: 1520-6106, 1520-5207. DOI: 10.1021/jp2086472.

[215] Hiroshi C. Watanabe et al. “Structural Model of Channelrhodopsin”. en. In: *Journal of Biological Chemistry* 287.10 (Mar. 2012), pp. 7456–7466. ISSN: 00219258. DOI: 10.1074/jbc.M111.320309.

[216] Yanan Guo et al. “Active site structure and absorption spectrum of channelrhodopsin-2 wild-type and C128T mutant”. en. In: *Chemical Science* 7.6 (2016), pp. 3879–3891. ISSN: 2041-6520, 2041-6539. DOI: 10.1039/C6SC00468G.

[217] Yanan Guo et al. “Different hydrogen bonding environments of the retinal protonated Schiff base control the photoisomerization in channelrhodopsin-2”. en. In: *Physical Chemistry Chemical Physics* 20.43 (2018), pp. 27501–27509. ISSN: 1463-9076, 1463-9084. DOI: 10.1039/C8CP05210G.

[218] Marius Wanko et al. “Effect of Polarization on the Opsin Shift in Rhodopsins. 1. A Combined QM/QM/MM Model for Bacteriorhodopsin and Pharaonis Sensory Rhodopsin II”. en. In: *The Journal of Physical Chemistry B* 112.37 (Sept. 2008), pp. 11462–11467. ISSN: 1520-6106, 1520-5207. DOI: 10.1021/jp802408g.

[219] Marius Wanko et al. “Effect of Polarization on the Opsin Shift in Rhodopsins. 2. Empirical Polarization Models for Proteins”. en. In: *The Journal of Physical Chemistry B* 112.37 (Sept. 2008), pp. 11468–11478. ISSN: 1520-6106, 1520-5207. DOI: 10.1021/jp802409k.

[220] Patrizia Mazzeo et al. “Electrostatic embedding machine learning for ground and excited state molecular dynamics of solvated molecules”. en. In: *Digital Discovery* 3.12 (2024), pp. 2560–2571. ISSN: 2635-098X. DOI: 10.1039/D4DD00295D.

[221] Elena Betti et al. “Insights into Energy Transfer in Light-Harvesting Complex II Through Machine-Learning Assisted Simulations”. In: *The Journal of Physical Chemistry B* 128.21 (May 2024). Publisher: American Chemical Society, pp. 5188–5200. ISSN: 1520-6106. DOI: 10.1021/acs.jpcb.4c01494.

[222] Julia Westermayr et al. “Deep learning study of tyrosine reveals that roaming can lead to photodamage”. en. In: *Nature Chemistry* 14.8 (Aug. 2022), pp. 914–919. ISSN: 1755-4330, 1755-4349. DOI: 10.1038/s41557-022-00950-z.

[223] J. Poštulka et al. “Diffusive versus non-diffusive paths to interstellar hydrogen peroxide: A machine-learning-based molecular-dynamics study”. en. In: *Astronomy & Astrophysics* 697 (May 2025), A51. ISSN: 0004-6361, 1432-0746. DOI: 10.1051/0004-6361/202452617.

[224] Julia Westermayr and Philipp Marquetand. “Machine Learning for Electronically Excited States of Molecules”. en. In: *Chemical Reviews* 121.16 (Aug. 2021), pp. 9873–9926. ISSN: 0009-2665, 1520-6890. DOI: 10.1021/acs.chemrev.0c00749.

[225] Julia Westermayr, Michael Gastegger, and Philipp Marquetand. “Combining SchNet and SHARC: The SchNarc Machine Learning Approach for Excited-State Dynamics”. en. In: *The Journal of Physical Chemistry Letters* 11.10 (May 2020), pp. 3828–3834. ISSN: 1948-7185, 1948-7185. DOI: 10.1021/acs.jpclett.0c00527.

[226] Sascha Mausenberger et al. “S pai NN: equivariant message passing for excited-state nonadiabatic molecular dynamics”. en. In: *Chemical Science* 15.38 (2024), pp. 15880–15890. ISSN: 2041-6520, 2041-6539. DOI: 10.1039/D4SC04164J.

[227] Monja Sokolov et al. “Non-adiabatic molecular dynamics simulations provide new insights into the exciton transfer in the Fenna–Matthews–Olson complex”. en. In: *Physical Chemistry Chemical Physics* 26.28 (2024), pp. 19469–19496. ISSN: 1463-9076, 1463-9084. DOI: 10.1039/D4CP02116A.

[228] Amanda Arcidiacono et al. “Predicting Solvatochromism of Chromophores in Proteins through QM/MM and Machine Learning”. en. In: *The Journal of Physical Chemistry A* 128.18 (May 2024), pp. 3646–3658. ISSN: 1089-5639, 1520-5215. DOI: 10.1021/acs.jpca.4c00249.

[229] Pavlo O. Dral et al. “Semiempirical Quantum-Chemical Orthogonalization-Corrected Methods: Benchmarks for Ground-State Properties”. en. In: *Journal of Chemical Theory and Computation* 12.3 (Mar. 2016), pp. 1097–1120. ISSN: 1549-9618, 1549-9626. DOI: 10.1021/acs.jctc.5b01047.

[230] Deniz Tuna et al. “Semiempirical Quantum-Chemical Orthogonalization-Corrected Methods: Benchmarks of Electronically Excited States”. In: *Journal of Chemical Theory and Computation* 12.9 (Sept. 2016). Publisher: American Chemical Society, pp. 4400–4422. ISSN: 1549-9618. DOI: 10.1021/acs.jctc.6b00403.

[231] Beatrix M. Bold et al. “Benchmark and performance of long-range corrected time-dependent density functional tight binding (LC-TD-DFTB) on rhodopsins and light-harvesting complexes”. en. In: *Physical Chemistry Chemical Physics* 22.19 (2020), pp. 10500–10518. ISSN: 1463-9076, 1463-9084. DOI: 10.1039/C9CP05753F.

[232] Mohsen M. T. El-Tahawy, Artur Nenov, and Marco Garavelli. “Photoelectrochromism in the Retinal Protonated Schiff Base Chromophore: Photoisomerization Speed and Selectivity under a Homogeneous Electric Field at Different Operational Regimes”. en. In: *Journal of Chemical Theory and Computation* 12.9 (Sept. 2016), pp. 4460–4475. ISSN: 1549-9618, 1549-9626. DOI: 10.1021/acs.jctc.6b00558.

[233] Pedro B. Coto et al. “The color of rhodopsins at the *ab initio* multiconfigurational perturbation theory resolution”. en. In: *Proceedings of the National Academy of Sciences* 103.46 (Nov. 2006), pp. 17154–17159. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.0604048103.

[234] Berk Hess et al. “GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation”. In: *Journal of Chemical Theory and Computation* 4.3 (Mar. 2008). Publisher: American Chemical Society, pp. 435–447. ISSN: 1549-9618. DOI: 10.1021/ct700301q.

[235] Carl Caleman et al. “Force Field Benchmark of Organic Liquids: Density, Enthalpy of Vaporization, Heat Capacities, Surface Tension, Isothermal Compressibility, Volumetric Expansion Coefficient, and Dielectric Constant”. In: *Journal of Chemical Theory and Computation* 8.1 (Jan. 2012). Publisher: American Chemical Society, pp. 61–74. ISSN: 1549-9618. DOI: 10.1021/ct200731v.

[236] Katharina Spies, Beatrix M. Bold, and Marcus Elstner. “Complex active site structures influence absorption spectrum of Chrimson wild type and mutants”. In: *Phys. Chem. Chem. Phys.* 27 (June 2025), pp. 13360–13370. ISSN: 1463-9084. DOI: 10.1039/D5CP00762C.

[237] Schrödinger, LLC. “The PyMOL Molecular Graphics System, Version 1.8”. Nov. 2015.

[238] Jing Huang and Alexander D. MacKerell. “CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data”. en. In: *Journal of Computational Chemistry* 34.25 (Sept. 2013), pp. 2135–2145. ISSN: 01928651. DOI: 10.1002/jcc.23354.

[239] Jeffery B. Klauda et al. “Update of the CHARMM All-Atom Additive Force Field for Lipids: Validation on Six Lipid Types”. In: *The Journal of Physical Chemistry B* 114.23 (June 2010). Publisher: American Chemical Society, pp. 7830–7843. ISSN: 1520-6106. DOI: 10.1021/jp101759q.

[240] M. Eichinger et al. “A hybrid method for solutes in complex solvents: Density functional theory combined with empirical force fields”. In: *The Journal of Chemical Physics* 110.21 (June 1999), pp. 10452–10467. ISSN: 0021-9606. DOI: 10.1063/1.479049.

[241] Hiroshi C. Watanabe et al. “Towards an Understanding of Channelrhodopsin Function: Simulations Lead to Novel Insights of the Channel Mechanism”. In: *Journal of Molecular Biology* 425.10 (May 2013), pp. 1795–1814. ISSN: 0022-2836. DOI: 10.1016/j.jmb.2013.01.033.

[242] Tino Wolter et al. “Mechanism by which Untwisting of Retinal Leads to Productive Bacteriorhodopsin Photocycle States”. In: *The Journal of Physical Chemistry B* 119.6 (Feb. 2015). Publisher: American Chemical Society, pp. 2229–2240. ISSN: 1520-6106. DOI: 10.1021/jp505818r.

[243] Julian J. Kranz and Marcus Elstner. “Simulation of Singlet Exciton Diffusion in Bulk Organic Materials”. en. In: *Journal of Chemical Theory and Computation* 12.9 (Sept. 2016), pp. 4209–4221. ISSN: 1549-9618, 1549-9626. DOI: 10.1021/acs.jctc.6b00235.

[244] M. Wanko et al. “Calculating Absorption Shifts for Retinal Proteins: Computational Challenges”. In: *The Journal of Physical Chemistry B* 109.8 (Mar. 2005). Publisher: American Chemical Society, pp. 3606–3615. ISSN: 1520-6106. DOI: 10.1021/jp0463060.

[245] Fabian Pedregosa et al. “Scikit-learn: Machine Learning in Python”. In: *Journal of Machine Learning Research* 12.85 (2011), pp. 2825–2830. ISSN: 1533-7928.

[246] Tim Sainburg, Leland McInnes, and Timothy Q. Gentner. “Parametric UMAP Embeddings for Representation and Semisupervised Learning”. In: *Neural Computation* 33.11 (Oct. 2021), pp. 2881–2907. ISSN: 0899-7667. DOI: 10.1162/neco_a_01434.

[247] Naveen Michaud-Agrawal et al. “MDAnalysis: A toolkit for the analysis of molecular dynamics simulations”. en. In: *Journal of Computational Chemistry* 32.10 (July 2011), pp. 2319–2327. ISSN: 0192-8651, 1096-987X. DOI: 10.1002/jcc.21787.

[248] Richard Gowers et al. “MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations”. en. In: Austin, Texas, 2016, pp. 98–105. DOI: 10.25080/Majora-629e541a-00e.

[249] Lisha Li et al. “Hyperband: A Novel Bandit-Based Approach to Hyperparameter Optimization”. en. In: ().

[250] *keras-team/keras-tuner*. original-date: 2019-06-06T22:38:21Z. Aug. 2025.

[251] Martın Abadi et al. “TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems”. en. In: ().

[252] Gerd G. Kochendoerfer et al. “Resonance Raman Examination of the Wavelength Regulation Mechanism in Human Visual Pigments”. en. In: *Biochemistry* 36.22 (June 1997), pp. 6577–6587. ISSN: 0006-2960, 1520-4995. DOI: 10.1021/bi970322o.

[253] Lars H. Andersen et al. “Absorption of Schiff-Base Retinal Chromophores in Vacuo”. en. In: *Journal of the American Chemical Society* 127.35 (Sept. 2005), pp. 12347–12350. ISSN: 0002-7863, 1520-5126. DOI: 10.1021/ja051638j.

[254] I. B. Nielsen, L. Lammich, and L. H. Andersen. “S 1 and S 2 Excited States of Gas-Phase Schiff-Base Retinal Chromophores”. en. In: *Physical Review Letters* 96.1 (Jan. 2006), p. 018304. ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.96.018304.

[255] Katharina Spies. *Computational Study of the Channelrhodopsin Chrimson Wild Type and Mutants*. Master Thesis. Apr. 2022.

[256] Karl Deisseroth and Peter Hegemann. “The form and function of channelrhodopsin”. en. In: *Science* 357.6356 (Sept. 2017), eaan5544. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.aan5544.

[257] Valentina Emiliani et al. “Optogenetics for light control of biological systems”. In: *Nature Reviews Methods Primers* 2 (July 2022), p. 55. DOI: 10.1038/s43586-022-00136-4.

[258] Feng Zhang et al. “The Microbial Opsin Family of Optogenetic Tools”. English. In: *Cell* 147.7 (Dec. 2011). Publisher: Elsevier, pp. 1446–1457. ISSN: 0092-8674, 1097-4172. DOI: 10.1016/j.cell.2011.12.004.

[259] Hartmann Harz and Peter Hegemann. “Rhodopsin-regulated calcium currents in Chlamydomonas”. In: *Nature* 351 (June 1991). ADS Bibcode: 1991Natur.351..489H, pp. 489–491. ISSN: 0028-0836. DOI: 10.1038/351489a0.

[260] Johannes Vierock et al. “Molecular determinants of proton selectivity and gating in the red-light activated channelrhodopsin Chrimson”. en. In: *Scientific Reports* 7.1 (Aug. 2017), p. 9928. ISSN: 2045-2322. DOI: 10.1038/s41598-017-09600-8.

[261] David Urmann et al. “Photochemical Properties of the Red-shifted Channelrhodopsin Chrimson”. en. In: *Photochemistry and Photobiology* 93.3 (May 2017), pp. 782–795. ISSN: 00318655. DOI: 10.1111/php.12741.

[262] Riccardo Palombo et al. “Retinal chromophore charge delocalization and confinement explain the extreme photophysics of Neorhodopsin”. en. In: *Nature Communications* 13.1 (Nov. 2022), p. 6652. ISSN: 2041-1723. DOI: 10.1038/s41467-022-33953-y.

[263] Oliver P. Ernst et al. “Photoactivation of Channelrhodopsin*”. In: *Journal of Biological Chemistry* 283.3 (Jan. 2008), pp. 1637–1643. ISSN: 0021-9258. DOI: 10.1074/jbc.M708039200.

[264] Hui Li, Andrew D. Robertson, and Jan H. Jensen. “Very fast empirical prediction and rationalization of protein pKa values”. en. In: *Proteins: Structure, Function, and Bioinformatics* 61.4 (2005). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/prot.20660, pp. 704–721. ISSN: 1097-0134. DOI: 10.1002/prot.20660.

[265] Delphine C. Bas, David M. Rogers, and Jan H. Jensen. “Very fast prediction and rationalization of pKa values for protein-ligand complexes”. eng. In: *Proteins* 73.3 (Nov. 2008), pp. 765–783. ISSN: 1097-0134. DOI: 10.1002/prot.22102.

[266] Mats H. M. Olsson et al. “PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions”. In: *Journal of Chemical Theory and Computation* 7.2 (Feb. 2011). Publisher: American Chemical Society, pp. 525–537. ISSN: 1549-9618. DOI: 10.1021/ct100578z.

[267] E G Alexov and M R Gunner. “Incorporating protein conformational flexibility into the calculation of pH-dependent protein properties.” In: *Biophysical Journal* 72.5 (May 1997), pp. 2075–2093. ISSN: 0006-3495.

[268] Roxana E Georgescu, Emil G Alexov, and Marilyn R Gunner. “Combining conformational flexibility and continuum electrostatics for calculating pK(a)s in proteins.” In: *Biophysical Journal* 83.4 (Oct. 2002), pp. 1731–1748. ISSN: 0006-3495.

[269] M. R. Gunner and E. Alexov. “A pragmatic approach to structure based calculation of coupled proton and electron transfer in proteins”. eng. In: *Biochimica Et Biophysica Acta* 1458.1 (May 2000), pp. 63–87. ISSN: 0006-3002. DOI: 10.1016/s0005-2728(00)00060-8.

[270] Junjun Mao, Karin Hauser, and M. R. Gunner. “How cytochromes with different folds control heme redox potentials”. eng. In: *Biochemistry* 42.33 (Aug. 2003), pp. 9829–9840. ISSN: 0006-2960. DOI: 10.1021/bi027288k.

[271] Yifan Song, Junjun Mao, and M. R. Gunner. “Calculation of proton transfers in Bacteriorhodopsin bR and M intermediates”. eng. In: *Biochemistry* 42.33 (Aug. 2003), pp. 9875–9888. ISSN: 0006-2960. DOI: 10.1021/bi034482d.

[272] K. Hauser, J. Mao, and M. R. Gunner. “pH dependence of heme electrochemistry in cytochromes investigated by multiconformation continuum electrostatic calculations”. eng. In: *Biopolymers* 74.1-2 (June 2004), pp. 51–54. ISSN: 0006-3525. DOI: 10.1002/bip.20042.

[273] Zhenyu Zhu and M. R. Gunner. “Energetics of quinone-dependent electron and proton transfers in Rhodobacter sphaeroides photosynthetic reaction centers”. eng. In: *Biochemistry* 44.1 (Jan. 2005), pp. 82–96. ISSN: 0006-2960. DOI: 10.1021/bi048348k.

[274] Kai Welke et al. “Color Tuning in Binding Pocket Models of the Chlamydomonas-Type Channelrhodopsins”. In: *The Journal of Physical Chemistry B* 115.50 (Dec. 2011). Publisher: American Chemical Society, pp. 15119–15128. ISSN: 1520-6106. DOI: 10.1021/jp2085457.

[275] Kai Welke et al. “QM/MM simulations of vibrational spectra of bacteriorhodopsin and channelrhodopsin-2”. en. In: *Physical Chemistry Chemical Physics* 15.18 (2013), p. 6651. ISSN: 1463-9076, 1463-9084. DOI: 10.1039/c3cp44181d.

[276] Paul Bauer, Berk Hess, and Erik Lindahl. “GROMACS 2022 Manual”. In: (Feb. 2022). DOI: 10.5281/zenodo.6103568.

[277] H. J. C. Berendsen, D. van der Spoel, and R. van Drunen. “GROMACS: A message-passing parallel molecular dynamics implementation”. In: *Computer Physics Communications* 91.1 (Sept. 1995), pp. 43–56. ISSN: 0010-4655. DOI: 10.1016/0010-4655(95)00042-E.

[278] Anders S. Christensen et al. “Semiempirical Quantum Mechanical Methods for Noncovalent Interactions for Chemical and Biochemical Applications”. en. In: *Chemical Reviews* 116.9 (May 2016), pp. 5301–5337. ISSN: 0009-2665, 1520-6890. DOI: 10.1021/acs.chemrev.5b00584.

[279] William Humphrey, Andrew Dalke, and Klaus Schulten. “VMD – Visual Molecular Dynamics”. In: *Journal of Molecular Graphics* 14 (1996), pp. 33–38.

[280] John Stone. “\em An Efficient Library for Parallel Ray Tracing and Animation”. MA thesis. Computer Science Department, University of Missouri-Rolla, Apr. 1998.

[281] Marius Wanko. “Optical Excitations in Biological Systems: Multiscale-Simulation Strategies and Applications to Rhodopsins”. en. PhD thesis.

[282] Nils Schieschke et al. “Geometry dependence of excitonic couplings and the consequences for configuration-space sampling”. en. In: *Journal of Computational Chemistry* 42.20 (July 2021), pp. 1402–1418. ISSN: 0192-8651, 1096-987X. DOI: 10.1002/jcc.26552.

[283] Monja Sokolov et al. “Analytical Time-Dependent Long-Range Corrected Density Functional Tight Binding (TD-LC-DFTB) Gradients in DFTB+: Implementation and Benchmark for Excited-State Geometries and Transition Energies”. en. In: *Journal of Chemical Theory and Computation* 17.4 (Apr. 2021), pp. 2266–2282. ISSN: 1549-9618, 1549-9626. DOI: 10.1021/acs.jctc.1c00095.

[284] Florian Weigend and Reinhart Ahlrichs. “Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy”. en. In: *Physical Chemistry Chemical Physics* 7.18 (2005), p. 3297. ISSN: 1463-9076, 1463-9084. DOI: 10.1039/b508541a.

[285] Frank Neese. “The ORCA program system”. en. In: *WIREs Computational Molecular Science* 2.1 (Jan. 2012), pp. 73–78. ISSN: 1759-0876, 1759-0884. DOI: 10.1002/wcms.81.

[286] Frank Neese. “Software update: The ORCA program system—Version 5.0”. en. In: *WIREs Computational Molecular Science* 12.5 (Sept. 2022), e1606. ISSN: 1759-0876, 1759-0884. DOI: 10.1002/wcms.1606.

[287] Hideaki E. Kato et al. “Crystal structure of the channelrhodopsin light-gated cation channel”. en. In: *Nature* 482.7385 (Feb. 2012), pp. 369–374. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/nature10870.

[288] Oleksandr Volkov et al. “Structural insights into ion conduction by channelrhodopsin 2”. en. In: *Science* 358.6366 (Nov. 2017), eaan8862. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.aan8862.

[289] Hai Li et al. “Role of a Helix B Lysine Residue in the Photoactive Site in Channelrhodopsins”. en. In: *Biophysical Journal* 106.8 (Apr. 2014), pp. 1607–1617. ISSN: 00063495. DOI: 10.1016/j.bpj.2014.03.002.

[290] Klaus Gerwert, Erik Freier, and Steffen Wolf. “The role of protein-bound water molecules in microbial rhodopsins”. In: *Biochimica et Biophysica Acta (BBA) - Bioenergetics*. Retinal Proteins 1837.5 (May 2014), pp. 606–613. ISSN: 0005-2728. DOI: 10.1016/j.bbabio.2013.09.006.

[291] Daniel Herschlag and Margaux M. Pinney. “Hydrogen Bonds: Simple after All?” en. In: *Biochemistry* 57.24 (June 2018), pp. 3338–3352. ISSN: 0006-2960, 1520-4995. DOI: 10.1021/acs.biochem.8b00217.

[292] E. Podoliak et al. “A subgroup of light-driven sodium pumps with an additional Schiff base counterion”. en. In: *Nature Communications* 15.1 (Apr. 2024), p. 3119. ISSN: 2041-1723. DOI: 10.1038/s41467-024-47469-0.

[293] Ziwei Pang et al. “Unravelling the mechanism of glucose binding in a protein-based fluorescence probe: molecular dynamics simulation with a tailor-made charge model”. en. In: *Physical Chemistry Chemical Physics* 24.4 (2022), pp. 2441–2453. ISSN: 1463-9076, 1463-9084. DOI: 10.1039/D1CP03733A.

List of Figures

List of Tables